Diversity of Coronaviruses in New World Bats

Bárbara dos Santos Simões

Dissertação para obtenção do

Grau de Mestre em Microbiologia Médica

Julho 2019

Diversity of Coronaviruses in New World Bats

Bárbara dos Santos Simões

Dissertação para obtenção do

Grau de Mestre em Microbiologia Médica

Orientadores:

Prof. Dr. Jan Felix Drexler (Charité - Universitätsmedizin Berlin)

Prof. Dr. Ricardo Parreira (IHMT/UNL)

Charité - Universitätsmedizin Berlin

Julho 2019

Acknowledgements

I would like to thank Professor Dr Jan Felix Drexler for the opportunity given to me to do my thesis in the Institute. For all the knowledge that he gave to me about the world of science and most important everything about Virology.

To Andrea Rasche for all the time spent teaching me and helping me during my thesis and of course for having the patience for all my questions. To Gustavo Goés for helping me in the beginning and kindly supply data.

To all my colleagues from the group, for giving me the opportunity to learn new things on different topics and the support and strength that they gave during my thesis.

To Juli Jules, Toby, Fabian, Tali, Sofia for making me laugh in the hard times and making me believe that I could make and especially to Juli Jules for the support and help that she gave to me and of course for have entered in my life. Ich liebe euch!

To my friends Alex and Seohee for everything. Thanks for making my life better in Berlin and for having you in my life. To my friends back in Portugal, Joana “Bimba”, Sofs, Marta and Mariana for the support and for believing in me. To Gabi, thank you for everything. For believing and pushing me to accomplish my dreams, for the kind and cold words when needed. Even far away you could make me feel that I was next to you. Thanks, dwarf! To Felix for listening and give important life advice, for the support, for making me laugh so much and for standing by my side in my most difficult times. Eintracht Frankfurt International!

To my family and my dog Dudu for the support and unconditional love that they gave me and the opportunity to study abroad and for all the thoughts that they taught me during my life.

Abstract

Coronaviruses comprise important human and animal pathogens. Six coronavirus species can cause human illness, probably all having a zoonotic origin. The prominent role of bats for coronavirus evolution was discovered in the aftermath of the 2003 SARS (Severe Acute Respiratory Syndrome) outbreak, where bats were identified as the zoonotic origin of this virus and a great diversity of coronavirus was identified in bats around the world. Indeed, bats present the largest genetic diversity of coronavirus compared to other hosts, making them the major natural host. The Neotropical region harbours a high variety of bat species presenting enormous potential for coronavirus studies. However, most of the studies regarding coronavirus diversity were conducted in Old World bats.

In this project, I studied the genetic diversity of coronavirus from about 1000 Neotropical bats. I conducted phylogenetic analyses from several coronavirus polymerase fragments and characterized two full genomes, comprising one alphacoronavirus from a Brazilian bat (*Phyllostomus discolor*) and one betacoronavirus from a Costa Rican bat (*Pteronotus parnellii*). The *Phyllostomus* bat alphacoronavirus might be an ancient relative of the human alphacoronaviruses 229E and NL63 and their bat-related coronaviruses.

The *Pteronotus* bat betacoronavirus is an ancient sister clade of the clade c betacoronavirus to which the human coronavirus MERS (Middle East Respiratory Syndrome) belongs and might correspond to an ancient root of the origins of MERS. In sum, this thesis expands the knowledge of coronavirus diversity in New World bats as and gives a deeper insight into the origins of the human coronavirus MERS, 229E and NL63.

Keywords: Coronavirus, Bats, Evolution, Phylogeny

Resumo

Os Coronavírus são patógenos importantes para o Homem e animais. Seis espécies de Coronavírus podem causar doenças no Homem e provavelmente têm uma origem zoonótica. Após a epidemia de SARS (*Severe Acute Respiratory Syndrome*), foi descoberto o papel proeminente dos morcegos para a evolução de coronavírus, na medida em que estes foram identificados como sendo a sua origem zoonótica. Após a epidemia, uma grande diversidade de coronavírus foi identificada em morcegos de todo o mundo. Estes apresentam a maior diversidade genética de coronavirus em comparação com outros hospedeiros, tornando-se os seus hospedeiros naturais. A região Neotropical contém uma elevada diversidade de espécies de morcegos, apresentando um enorme potencial para estudos de diversidade de coronavírus contudo, a maioria destes estudos só foi realizado com morcegos do Velho Mundo.

Neste projeto, eu estudei a diversidade genética de coronavírus em aproximadamente 1000 morcegos do Neotrópico. Eu realizei análises filogenéticas de vários fragmentos da polimerase de coronavírus e caracterizei dois genomas completos: um alphacoronavírus presente num morcego do Brazil (*Phyllostomus discolor*) e um betacoronavírus de um morcego da Costa Rica (*Pteronotus parnellii*). O alphacoronavírus pode ser um parente antigo deste grupo e pode estar relacionado com os alphacoronavírus humanos 229E e NL63 e 229E e NL63 relacionados com morcegos.

O betacoronavírus também pertence a um antigo grupo relacionado com o grupo C em que o coronavírus humano MERS (*Middle East Respiratory Syndrome*) pertence. Isto pode corresponder a uma possível explicação para a origem de MERS. Em resumo, esta tese contribui para o aprofundamento dos conhecimentos da diversidade de coronavírus em morcegos do Novo Mundo e fornece uma visão mais aprofundada sobre as origens dos coronavírus humanos MERS, 229E e NL63.

Palavras-chave: Coronavírus, Morcegos, Evolução, Filogenia

Index

Figure index

Table index

Abbreviations

°C	*Celsius* degrees
3Clpro	3C-like main protease
µL	Microliter= 10^{-6} Liter
a.a	Amino acid
ADRP	ADP-ribose 1-phosphatase
BCoV	Bovine coronavirus
BEAST	Cross-platform program; Bayesian Evolutionary Analysis Sampling Trees
BLAST	Bioinformatic tool; Basic Local Alignment Search Tool
bp	basepairs
BSA	Bovine Serum Albumi
CoV	Coronavirus
DNA	Desoxyribonucleic acid
dNTP	2'-Desoxyribonucleoside-5'-triphosphate
E	Envelope Protein
et al.	From Latin: *et alii/et aliae* (meaning: and others)
ExoN	3'-to-5' exoribonuclease
FIPV	Feline Infectious Peritonitis Virus
HE	Hemagglutinin-esterase
HEL	Helicase
Hz	Hertz
IBV	Infectious Bronchitis Virus
ICTV	International Committee for the Taxonomy of Viruses
IFN	Interferon

M	Membrane glycoprotein
MAFFT	Multiple Alignment program; Multiple Alignment using fast Fourier transform
MERS-CoV	Middle East Respiratory Syndrome coronavirus
Mg	milligram=10^{-3} Kilo
$MgCl2_2$	Magnesium Chlorate
$MgSO4_4$	Magnesium Sulfate
Min	Minute
mM	Milimole=10^{-3} molar
mRNA	Messenger Ribonucleic Acid
N	Nucleocapsid Protein
NCBI	National Center for Biotechnology Information
NendoU	Nidoviral uridylate-specific endoribonuclease
NMT	N7 methyltransferase
No.	Number
NSP	Nonstructural protein
Nt	Nucleotide
NTPase	Nucleoside triphosphatase
O-MT	Ribose-2'-O-methyltransferase
ORF	Open Reading Frame
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PEDV	Porcine Epidemic Diarrhea Virus
Phyre	Bioinformatic tool; Protein Homology/analogY Recognition Engine

Plpro	Papain-like protease
Sec	Second
SADS	Swine Acute Diarrhea Syndrome
SARS-CoV	Severe Acute Respiratory Syndrome coronavirus
SSE	Sequence Editor, Database and Analysis Platform; Simple Sequence editor
RBD	Receptor-binding domain
RDP4	Recombination Detection Program
RdRp	RNA-dependent RNA polymerase
RGU	RdRp Grouping Units
RNA	Ribonucleic Acid
RT-PCR	Reverse transcriptase Polymerase Chain Reaction
S	Spike glycoprotein
SARS	Severe Acute Respiratory Syndrome
ss	Single-Stranded
TBE	TRIS-borate
TGEV	Transmissible Gastroenteritis Virus of Swine
Tm	Melting Temperature. The temperature at which half of the DNA strand is in the random coil or single-stranded state
TRS	Transcription regulatory sequence
TRS-B	Body transcription regulatory sequence
TRS-L	Leader transcription regulatory sequence

UV	Ultraviolet
V	Volts
WAG	Whelan and Goldman

1. Introduction

1.1. Family *Coronaviridae*

1.1.1. Taxonomy of coronaviruses

The order Nidovirales is divided into 3 families: *Arteriviridae*, *Roniviridae* and *Coronaviridae.* This last family is divided into 2 subfamilies: *Coronavirinae* and *Torovirinae*. *Coronaviridae* comes from the Latin word "Corona" which means "halo" or "crown" and refers to the appearance of projections on the surface that resemble a solar corona. This family, commonly known as coronavirus (CoV) is a monophyletic family of the order Nidovirales and includes four genera: *alpha-, beta-*, *gamma-* and *deltacoronavirus*. In addition, the *betacoronavirus* genus is furthermore divided into 4 clades: A to D (Figure 1) (De Groot et al., 2012).

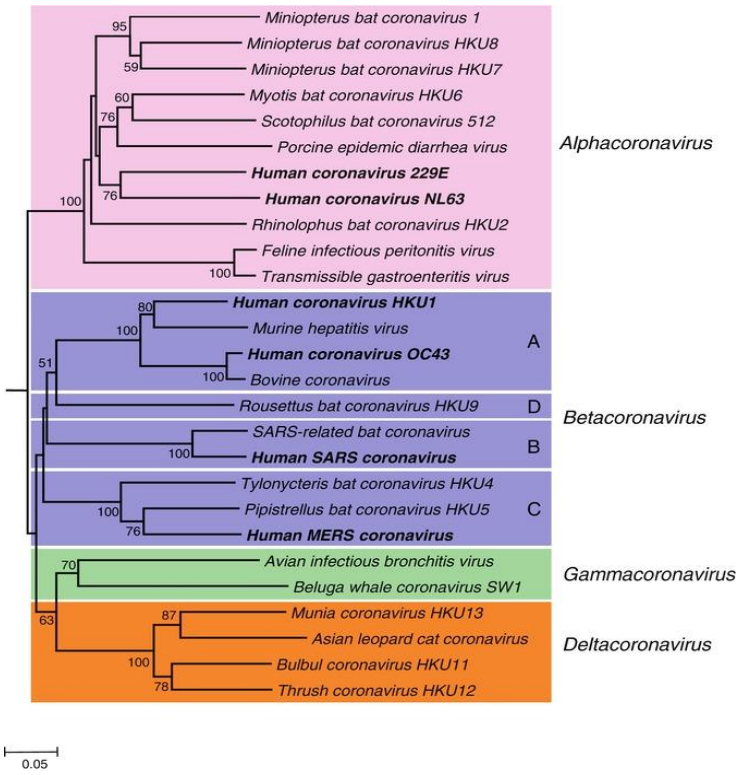


Figure 1 - Diversity of CoV (Kaslow et al., 2014).

The genera *Alpha-* and *Betacoronavirus* mainly infect mammals while the genera *Gamma-* and *Deltacoronavirus* infect mainly birds. However, members of the genus *Deltacoronavirus* were also found in pigs and in Asian leopards and members of the genus *Gammacoronavirus* were found in the beluga whale and in other cetaceans (Drexler et al., 2014; Woo et al., 2014; Woo et al., 2012).

CoVs are usually related to respiratory, enteric and hepatic to central nervous system illness depending on the virus clade and host, in both birds and mammals including humans (Lai & C., 2007; Weiss & Navas-Martin, 2005).

1.1.2. Morphology and genome organization

CoVs possess a positive single-stranded, nonsegmented genome. Described as a large RNA virus, the genome can range between 26-32 kb with a 5' terminal cap structure and a 3' poly-A tail. Almost two-thirds of the genome holds 2 Open Reading Frames (ORFs): ORF1a and ORF1b which encode in total 16 non-structural proteins (NSP). The polyprotein pp1a encodes for the NSPs1 to 10 and pp1b for NSP12 to 16, which includes major proteins such as the RNA- dependent RNA polymerase (RdRp) (NSP12) and Helicase (HEL) (NSP13). The rest of the genome encodes for essential structural proteins such as Envelope, Spike, Membrane, Nucleocapsid and accessory proteins (Lai & Cavanagh, 1997) (Figure 2). The accessory proteins are frequently interspaced between the major ORFs and their number varies with a species and most of these proteins do not play a role in virus replication cycle. However, it has been demonstrated that some accessory proteins are important for the virus pathogenicity, like the Murine CoV ns2 protein (Zhao et al., 2012). These proteins were initially labeled as nonstructural proteins but later it was demonstrated that some proteins are components of the virion like for example the Hemagglutinin-esterase (HE) (Fehr & Perlman, 2015; Masters, 2006).

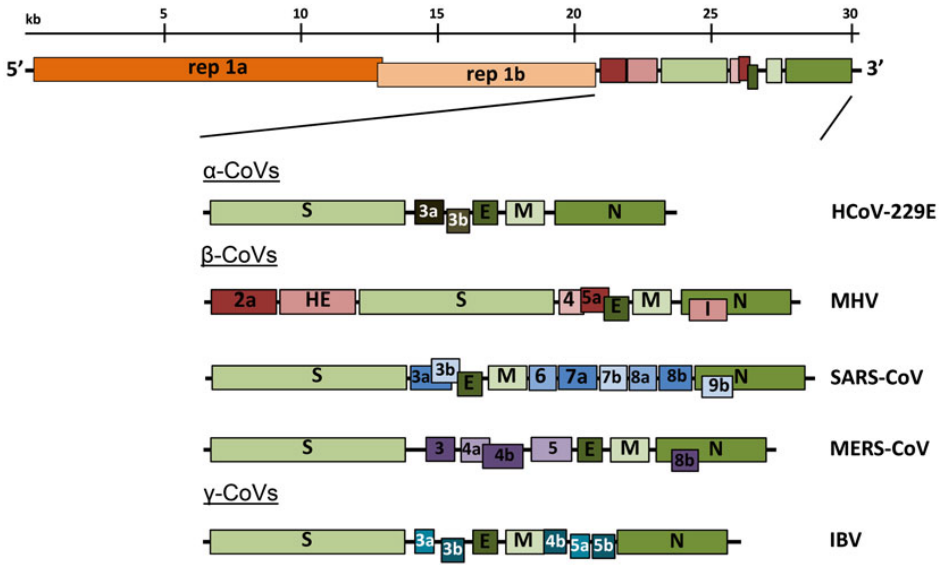


Figure 2 - Schematic representation of the genome organization of the main Open Reading Frames (ORFs) of important CoV. Top to bottom: Human alphacoronavirus 229E (hCoV-229E); betacoronavirus: Murine hepatitis virus (MHV), Severe acute respiratory syndrome (SARS-CoV), Middle East respiratory syndrome (MERS-CoV); gammacoronavirus Infectious bronchitis virus (IBV) (Fehr & Perlman, 2015).

The virion holds 4 structural proteins: Nucleocapsid (N) protein, Membrane (M) glycoprotein, Spike (S) glycoprotein, Small envelope protein (E) and in some cases, CoV like Betacoronavirus 1, Murine CoV and human CoV (hCoV) HKU11 (clade A betacoronavirus) possesses hemagglutinin-esterase (Figure 3).

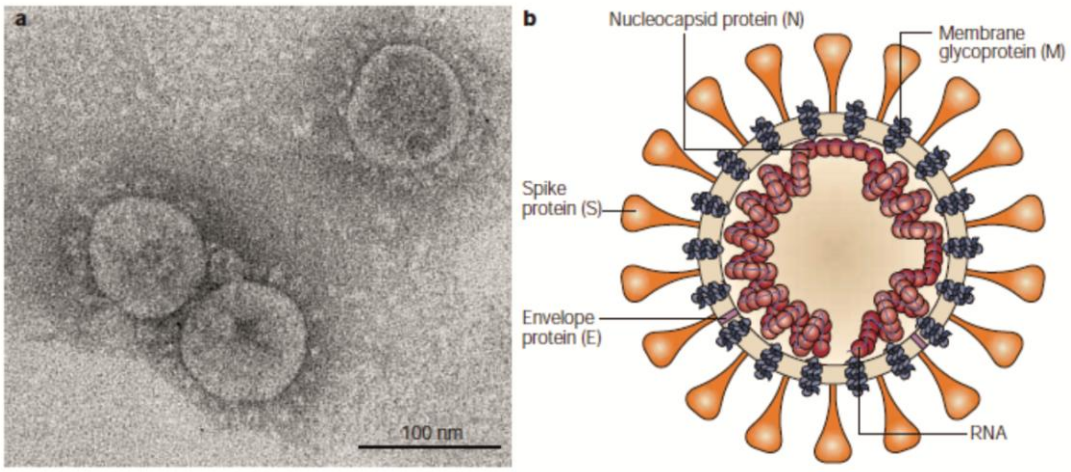


Figure 3 - CoV morphology. (a) Morphology of SARS-CoV observed by electron microscopy. (b) Schematic representation of a CoV virion: E, Envelope protein; N, Nucleocapsid protein; S, Spike protein; M, Membrane glycoprotein (Stadler et al., 2003).

The Spike glycoprotein accounts for the entry and fusion by recognition of the cell receptors, an important factor for the determination of viral host tropism. It contains 3 structural domains between the N-terminus and the C-terminus and 2 subunits. The S2 subunit is responsible for the membrane-fusion while the S1 receptor binding subunit, which contains the receptor-binding domain (RBD) is responsible for the recognition of cell receptors (Gallagher & Buchmeier, 2001). The membrane protein is the most abundant protein in the CoV virion and it is responsible for giving shape to the envelope structure. This triple spanning transmembrane protein has a short amino-terminal domain in the exterior of the virion and a long carboxy-terminal domain in the inside (Rottier, 1995). Interactions between M proteins allows excluding some host proteins from the host membrane of the viral envelope (de Haan et al., 2000; Neuman et al., 2008).

In the nucleocapsid, the N protein is an RNA binding phosphoprotein and it is bonded to the viral RNA. This protein is divided into three conserved domains, with hypervariable regions between these conserved regions and plays a role in the genome encapsidation, RNA synthesis and translation and also acts as a type I interferon antagonist (De Groot et al., 2012). The E protein is a small polypeptide and it is a minor constituent of the virion envelope. This integral membrane protein with an ion channel and/or membrane permeabilizing activities plays a role in virion assembly and it has been identified as a virulence factor for SARS-CoV.

Also, the N protein can have viroporin activities playing an important role in the virus progeny production and might also interfere in host cellular functions, contributing to viral pathogenicity (De Groot et al., 2012; Masters, 2006; Nieva & Carrasco, 2015).

Some CoVs such as Betacoronavirus 1, Murine CoV and hCoV HKU11, hold in their envelope an additional glycoprotein: hemagglutinin-esterase (HE). This protein possesses a sialate-O-acetylesterase receptor destroying enzyme (RDE) activity, which acting as a lectin or a sialate-O-acetylesterase allows the virion to reversibly attach to O-acetylated sialic acids (Pfleiderer et al., 1991; Schultze et al., 1991; Vlasak et al., 1988).

1.1.3. Replication Cycle

After the attachment of the virus to the host membrane and the release of the viral RNA into the host cytoplasm, the replicase gene that encodes the 2 major ORFs, ORF1a and 1b is directly translated by the host ribosomes (Sola et al., 2011) (Figure 4). To accomplish the correct translation, the virus appeals to a ribosomal frameshifting mechanism. Within the ORF1ab there is a slippery sequence (5'-UUUAAAC-3') and an RNA pseudoknot which leads to a ribosomal stop of the elongation within the slippery sequence. This block makes the ribosome to move back one nucleotide and therefore changing the reading frame to -1 frameshift. Upon this, the host ribosome can continue the elongation of the rest of the ORF1ab and after the translation step, the poplyprotein1ab is cleaved into individual non-structural proteins (NSPs) (Fehr & Perlman, 2015; Lim et al., 2016).

After the translation of the major NSPs, the virus RNA is transcribed by the RdRp into an antisensenegative-strand RNA. This is later used for the generation of positive-strand RNA molecules and sub-genomic mRNA which is used to generate several structural and accessory proteins (Fehr & Perlman, 2015).

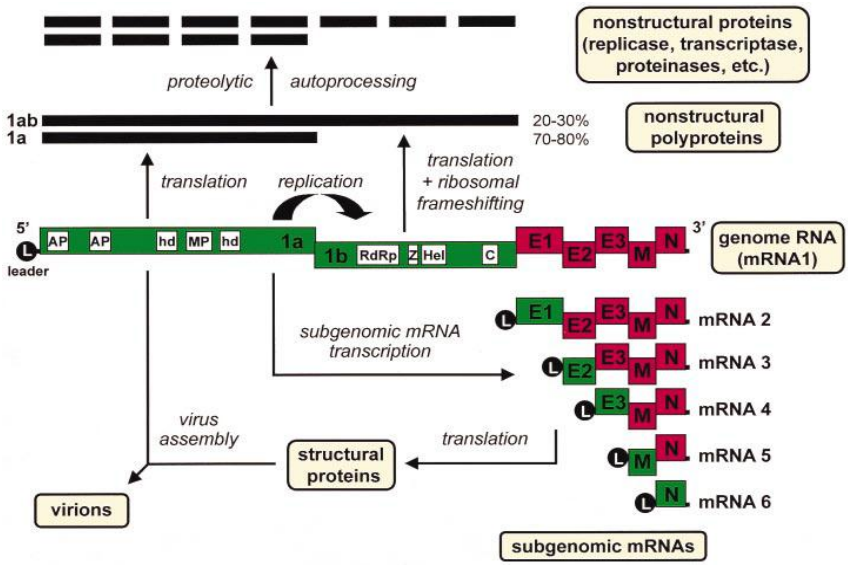


Figure 4 - Replication cycle of CoV (Gorbalenya et al., 2015).

The generation of sub-genomic RNA is held by a discontinuous transcription mechanism (figure 5). At the 5' end of the CoV genome and at the end of each mRNA coding sequence, there is a transcription regulatory sequence (TRS) which helps to mark the location of each mRNA coding sequence. The RdRp starts the replication in the 3' end of the positive-strand RNA and pauses at every TRS which in this case, is called a body TRS (TRS-B). When the RdRp encounters the TRS-B there are 2 options: the first option is to continue the elongation until it finds the next TRS-B or amplifies the leader TRS (TRS-L) located at the 5' end of the genome adding it to the 3' end of the newly synthesized strand. This process is repeated for all genes that code for the structural and accessory proteins. Following the translation of the structural and accessory proteins, the new virion is assembled and released into the host plasma (Fehr & Perlman, 2015; Sawicki et al., 2007; Sola et al., 2011).

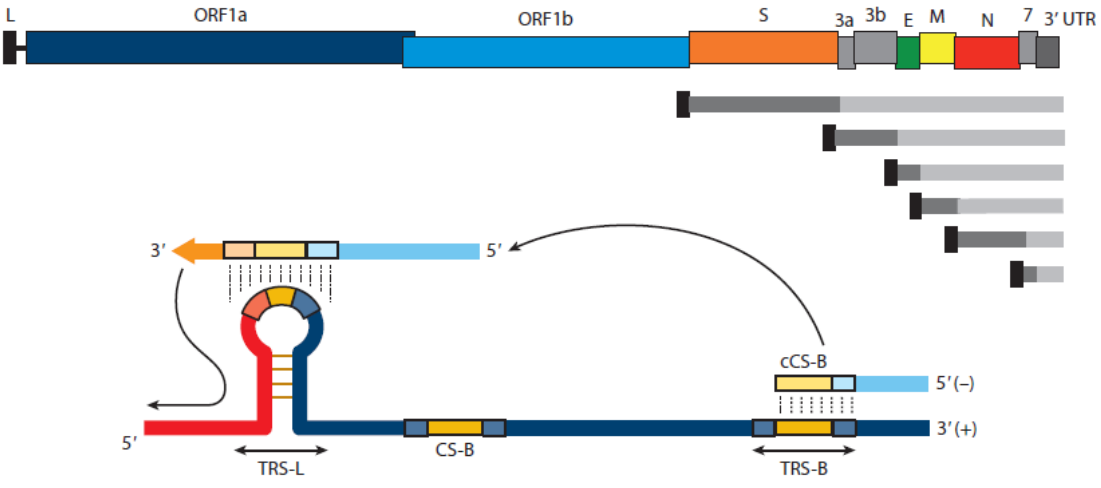


Figure 5 - Scheme of the discontinuous transcription of CoV. The short lines in the upper figure represent the nested set of sub-genomic mRNAs and the dark part represents the TRS-L. The dark grey area represents the gene to be translated. In the lower figure: CS-L, core sequence in the TRS-L; CS-B, core sequence of the TRS-B (Sola et al., 2015).

The classification of CoVs by the International Committee for the Taxonomy of Viruses (ICTV) analyses the amino acids (a.a) encoded by the following set of NSPs allowing the classification of a new CoV species: NSP3, NSP5, NSP12, NSP13, NSP14, NSP1 and NSP16. If the a.a sequence identity within these concatenated NSPs is inferior to 90% compared to species previously identified, a new species is classified. If the a.a similarity is inferior of 46% with the previous genera identified, a new genus is assigned. This method can have some problems such as the type of samples used like faeces which contain substances that contribute to RT-PCR assay inhibition making it difficult to sequence the NSPs and consequently, the full CoV genome to assign a new CoV species (Drexler et al., 2010).

Therefore, Drexler et. al., (2010) suggested an alternative method for the classification of CoV based on RdRp - grouping units (RGU), which is based on the comparison of 816 a.a sequence differences of the RdRp.

1.1.4. Diversity and evolution of coronaviruses

Within the 4 CoV genera, bats have been assigned as the ancestors of alpha and betacoronavirus whereas birds have been assigned with the ancestors of gamma and deltacoronavirus (Wong et al., 2019). A vast variety of pathogenic CoV have an evolutionary history of cross-species transmission events, as exemplified by several betacoronavirus such as SARS-CoV (clade B), MERS-CoV (clade C) and hCoV OC43 (clade A); and alphacoronaviruses such as: Porcine Epidemic Diarrhea Virus (PEDV), hCoVs 229E and NL63, Transmissible Gastroenteritis Virus of Swine (TGEV) and Feline Infectious Peritonitis Virus (FIPV) (Saif, 2004).

Genetic and evolutionary characteristics such as the large genome and high mutation rate of approximately 10^{-4} substitution per year per site give to this viral family a high level of plasticity. All these characteristics can help to explain the ability of adaptation to a new host and ecological niches (Woo et al., 2009; Woo et al., 2006).

The detection and importance of CoV were reinforced after the outbreaks caused by emerging CoVs. In 2003, a new epidemic emerged in the "wet markets" (type of market where is sold fresh fish and meat and where they also have living animals) in the Guangdong province, South of China (Guan et al., 2003; Ksiazek et al., 2003; Lai & C., 2007; Xu et al., 2004). A novel virus named SARS-CoV (Severe Acute Respiratory Syndrome) was discovered. Epidemiological and phylogenetic studies revealed the origin of this virus and indicated a bat origin, with the participation of civets (*Paguma larvata*) as an intermediated host.

However, direct transmission of CoV between bats and humans cannot be discarded. Since this epidemic, bats have been associated with CoVs.

In 2012, in the Middle East, a new CoV emerged. MERS-CoV (Middle East Respiratory Syndrome) was identified as the cause of the epidemic and its transmission from dromedaries to humans was probably due to multiple and frequents events of cross-species transmission (Corman et al., 2014).

Other emerging CoVs also have a big importance for humans and livestock. The hCoV-NL63, for example, presents a close relationship with Kenyan bats indicating a possible bat zoonotic origin, but the intermediate host origin it is yet not known.

Theories indicate a possible cross-species transmission event from ancient bat CoVs allowing them the ability to infect and cause diseases to humans (Corman et al., 2018; Tao et al., 2017). Phylogenetic studies indicate that hCoV-229E also has a zoonotic origin.

The identification of 229E-related CoVs in bats from Africa, alpacas from South America and dromedaries from South Arabia indicate a zoonotic origin and possibly a route of cross-species transmissions which led to the infection and stabilization in the human host. Other hCoVs like OC43 and HKU1 are hosted in cattle and for example, hCoV OC43 presents a high similarity with the Bovine CoV (BCoV), suggesting a possible origin from this BCoV (Corman et al., 2018; Lim et al., 2016).

Recently, in 2018 a new emerging CoV, SADS (Swine Acute Diarrhea Syndrome) that killed approximately 25.000 piglets in China, has been connected with the *Rhinolophus* bat genera. This emerging CoV was related, with high similarity to HKU2 bat CoV, being this HKU2 like CoV later nominated as SADS (Gong et al., 2017; Zhou et al., 2018). The emergence of this new virus indicates once more an event of cross-species transmission between bats and new hosts, and points out the necessity of surveillance of new events like this, in order to avoid in the future economic, human health and environmental setbacks.

Also, small insectivorous mammals like the hedgehog and shrew have been associated with CoVs. These 2 mammals belong to the order Eulipotyphla which also includes other small insectivorous mammals such as moles and solenodons. The order Eulipotyphla is phylogenetically related to the Chiroptera order to which bats belong (Bininda-Emonds et al., 2007; Corman et al., 2014).

In sum, bats are important hosts for CoVs. Their ability to fly and migrate as well as the dimension of their social groups makes them an important host for the acquisition and maintenance of viruses (Moratelli & Calisher, 2015).

1.2. Neotropical bats

1.2.1. New and Old World bats

Bats emerged about 64 million years ago in the Cretaceous period (figure 6) (Teeling et al., 2005). They constitute approximately 20% of mammals species and are distributed through all continents except Antarctica, with more than 350 species divided into 3 superfamilies (Emballonuroidea, Noctilionoidea and Vespertilionoidea) (López-Aguirre et al., 2018; Simmons, 2005).

The Neotropical bats emerged about approximately 30 million years ago (Teeling et al., 2005). In the Neotropics (Biographic region that includes South and Central America, southern part of Florida and Mexico, as well as West Indies) out of 18 bats families that exist in the world, only six families are endemic in the region and only 3 occur in both the New and Old World as shown in Figure 6. The Neotropical bats are present in a vast number of ecological niches and have a wide feeding habits from insectivorous, frugivorous to nectarivorous and carnivorous (Masters, 2006; Rex et al., 2008; Simmons, 2005; Teeling et al., 2005).

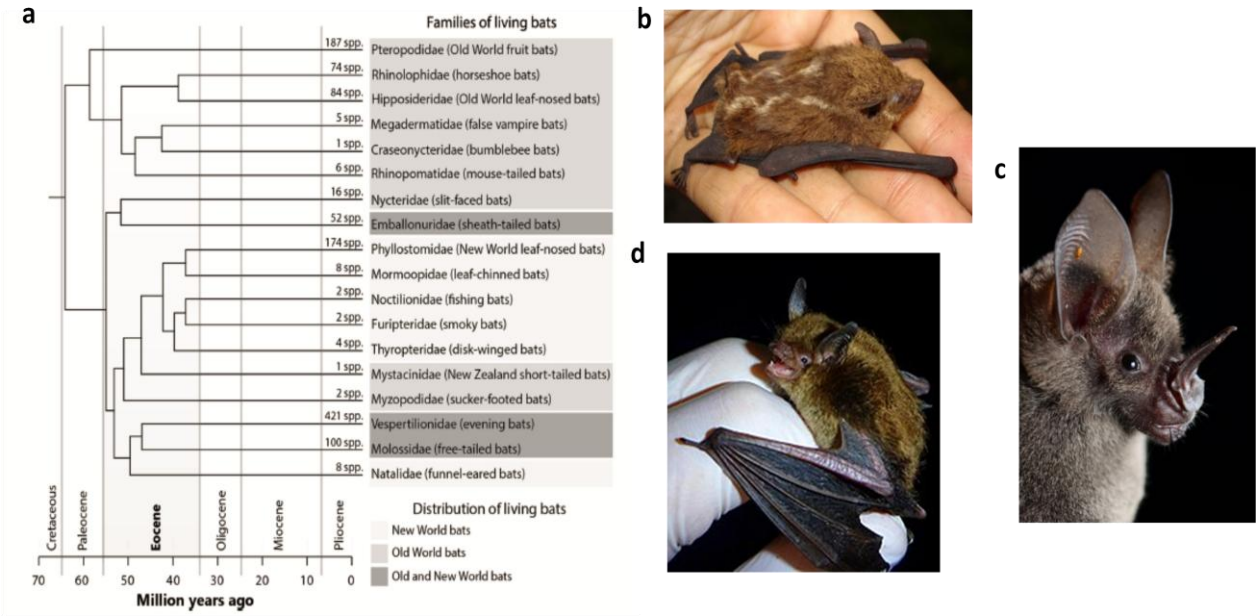


Figure 6 - Phylogeny of bats. (a) The time of the bat origin is represented in million years ago and described with the geological periods (Peixoto et al., 2018). (b) Emballonuroidae superfamily. In the image is represented the Lesser Sac-winged bat. Access date and link: 16.07.2019 https://stricollections.org /portal/taxa/index.php?tid=54181&taxauthid=1&clid=50. (c) Noctilionoidea superfamily represented by Lesser spear-nosed bat. Access date and link: 16.07.2019 https://www.inaturalist.org/taxa/41080-Phyllostomus-elongatus. d) Vespertilionoidea superfamily represented by Little brown bat. Access date and link: 16.07.2019 https://en.wikipedia.org/wiki /Little_brown_bat.

Bats are natural reservoirs for a vast number of viruses. For over 100 years, bats have associated with rabies virus and later with other human viruses, such as Hendravirus, Nipahvirus, Paramyxovirus, Filoviruses and CoV (Calisher et al., 2006). In the aftermath of the SARS- CoV epidemic, in particular, several bat CoVs have been described (Calisher et al., 2006; Chen et al., 2014). However, the number of studies yield in the Neotropical bats is very scarce and only partial sequence fragments have been acquired in all studies harboured in the Neotropics. Until now only a few studies were performed on the Neotropical bats. One was performed in bats from Trinidad and Tobago where two alphacoronavirus were described (Carrington et al., 2008), another one from Mexican bats (Anthony et al., 2013) where both beta- and alphacoronavirus clustered in 13 different clades. Other 2 studies were conducted in bats from Costa Rica, one which 5 alpha and 2 betacoronavirus were described (Corman et al., 2013) and another where several alphacoronaviruses were described (Moreira-Soto et al., 2015).

In addition, since 2009 was established a Consortium designed to address the global diversity of CoVs and over the years it has acquired partial genomes from bat CoVs from different parts of the world such as Africa, Latin America and Asia (Anthony et al., 2017; Consortium, 2014).

1.3. Aim of the Thesis

This thesis with the purpose of obtaining a Master degree in Medical Microbiology had as main goals the assessment of the CoVs diversity in South America throughout the analyzes of samples from Brazil. In addition, 2 full genomes were analyzed and characterized in order to examine the diversity of CoV from South America. One unpublished full genome from a Costa Rica *Pteronotus parnellii* bat CoV and another one from a Brazilian *Phyllostomus discolor* bat CoV obtained during the study of CoVs' diversity.

2. Materials and Methods

2.1. Bat sampling and processing

2.1.1. Origin and capture of bat samples

In a previous study (Goés et. al. 2016), 1004 bat samples (n=987 Intestines; n=15 rectal swabs and n=2 serum) were sampled in 4 different locations of Brazil. Out of the 1004 bat samples, 32 samples were previously screened for CoV in (Góes et al., 2016). All sampling, capture and sample transfers were done with the proper wildlife permits and ethics clearance and complied with the current laws of host countries. The samples were transported from Brazil to the Institute for Virology-Charité (Berlin, Germany) in styrofoam boxes with dry ice to prevent the samples to thaw. After receiving the samples they were stored at -80°C before further study.

2.2. Sample extraction

From 972 intestine samples that were not previously screened, approximately 30 mg was cut with the help of a sterile scalpel, Petri dishes and inside a hood, to prevent contamination of the outside of the hood and the samples. The samples were placed in a safety lock Eppendorf with 500 µL of PBS (Phosphate-Buffered Saline) and a metallic bead. The samples were homogenized using the TissueLyser II equipment (Qiagen, Hilden, Germany) at 30 Hz for 3 min. After the samples were centrifuged, pools of 8 samples were prepared whereas 35 µL of each sample was transferred to a new tube.

The pools were mixed and centrifuged using a vortex and a centrifuge, and 250 µL of the homogenate was transferred to a plate with 250 µL of Tissue Lyser buffer. The purification of the pools was performed in the MagNA Pure 96 System (Roche, Prenzberg, Germany) with elution of 100 µL. The intestine pools were purified using a Large volume kit (Roche, Prenzberg, Germany) and the swab and serum already screened using a Small volume kit (Roche, Prenzberg, Germany).

2.3. Screening assays for coronavirus

The CoV screening was divided into two parts more specifically the screening for betacoronavirus and alphacoronavirus due to the high variability and diversity among this 2 different genus.

Regarding the alphacoronavirus screening assay, all screening assays available (Anindita et al., 2015; Anthony et al., 2017; Chu et al., 2011; de Souza Luna et al., 2007; Falcón et al., 2011; Hu et al., 2018; Lacroix

et al., 2017; Lelli et al., 2013; Moreira-Soto et al., 2015; Quan et al., 2010; Vijgen et al., 2008; Wacharapluesadee et al., 2015; Watanabe et al., 2010) and other primers present in the Institute were evaluated using a bioinformatics approach in order to verify which assay could properly cover all diversity of alphacoronavirus. All ICTV reference sequences from the 4 CoV genera (alpha-, beta-, gamma- and deltacoronavirus) and other relevant alphacoronavirus sequences, from which complete genomes were available, were aligned using MAFFT alignment, on Geneious Prime software. The primers from each screening assay were annotated in the sequences and the primer binding regions and their mismatches were compared to all reference sequences and alphacoronavirus sequences.

The presence of betacoronavirus RNA was detected by using a heminested RT PCR (*Reverse transcription PCR*). A 228nt fragment from the RdRp gene was amplified using the following primers (5'-3') (Annan et al., 2013) : Pan2cRdRP-R: GCATWGCNCWGTCACACTTAGG; Pan2cRdRp-Rnest: CACTTAGGRTARTCC CAWCCA; Pan2cRdRp-FWD: TGCTATWAGTGCTAAGAATAGRGC.

For the first PCR round and for each reaction, 2,5 µL of the purified RNA was used as a template and added to a master mix constituted by 1,55 µL of RNase free water; 6,25 µL of 2X Reaction Mix (buffer containing 0,4 mM of each dNTP; 3,2 mM of $MgSO_4$); 0,2 µL of $MgSO_4$ (50mM); 0,5 µL BSA (Bovine Serum Albumin) (1mg/mL); 0,5 µL of each Pan2cRdRp-FWD and Pan2cRdRP-R primers and 0,5 µL of Superscript III RT Mix. For the second PCR round and for each reaction, 1 µL first round product was added to the following master mix constitution: 17,15 µL of RNase free water; 2,5 µL of 10X PCR Buffer without $MgCl_2$; 0,5 µL of dNTP Mix (10 mM of each dNTP); 1,25 µL of $MgCl_2$ (50 mM); 1 µL of Pan2cRdRp-FWD; 1,5 µL of Pan2cRdRP-Rnest and 0,1 µL of Platinium *Taq* DNA Polymerase. The cycling protocols used for the first and second steps of the heminested RT-PCR are shown in table 1 and 2.

Table 1- Cycling protocol of the first round of the betacoronavirus screening assay.

Step	Temperature (°C)	Duration	Number of cycles
Reverse transcription	50°C	20 min	
Initial denaturation	95°C	3 min	
Denaturation	94°C	15 sec	
Annealing	60°C*	15 sec	20X
Elongation	72°C	30 sec	
Denaturantion	95°C	15 sec	
Annealing	50°C	15 sec	30X
Elongation	72°C	30 sec	
Final Elongation	72°C	1 min	

*Touchdown of -0.5 °C per cycle. Temperature decreases 0.5°C per cycle.

Table 2- Cycling protocol of the second round of the betacoronavirus screening assay.

Step	Temperature (°C)	Duration	Number of cycles
Initial denaturation	95°C	3 min	
Denaturation	94°C	15 sec	20X
Annealing	60°C*	15 sec	
Elongation	72°C	30 sec	
Denaturation	95°C	15 sec	30X
Annealing	50°C	15 sec	
Elongation	72°C	30 sec	
Final Elongation	72°C	1 min	

*Touchdown of -0.5 °C per cycle.

The second round PCR product was analyzed and separated by gel electrophoresis. The samples were loaded into agarose gels. 2% agarose gels were done using 1x TBE buffer (TRIS-*borate*) and added 5 µL of Midori Green Advance DNA staining. 3 µL of the PCR product was mixed with in 2 µL of 5x gel loading dye pH 7.0. To determine the size of each band, a DNA Ladder was used as a standard reference. The gels were placed in a chamber with 1X TBE buffer and the run was done with 230V during approximately 20min. The gel with the separated bands was observed under UV-light. The PCR products that had the correct amplicon size band were sent for sequencing as explained further in section 2.5 of the Materials and Methods.

2.4. Extension of PCR product with specific primers

The 394nt screening fragment of the positive samples which were previously screened for CoV was extended to achieve the 816 nt fragment of the RdRp. Based on sequence similarities, samples were divided into 4 groups: Group I (G1) Alpha-CoV, Group II (G2) Alpha-CoV, Sturnira and Beta-CoV. For each group reverse primers were designed specifically. All primers used are listed in table 3 and were annotated in aligned sequences of the screening fragments and full genomes of the closely related CoV. The primers were verified on oligocalc online software (http://biotools.nubic.northwestern.edu/OligoCalc.html) to analyze the melting temperature and self-complementary properties.

Table 3 - List of the primers used for the elongation of the screening fragment.

Name	Sequence (5'-3')	Assay	Tm
AlphaCoVG1 RGU R1 (10mM)	TTGCAMARACACTARTACYTCTCACT	Hemi-Nested G1 AlphaCoV; 1st round	56,8°C-67,9°C
AlphaCoVG1 RGU R2 (10mM)	GAACCMAGWATCATGGCDGAGATCAT	Hemi-Nested G1 AlphaCoV; 2nd round	64,6°C-67,9°C
AlphaCoVG2 RGU R1 (10mM)	TAGAAWCCACCATTDGARTAMACAAC	Hemi-Nested G2 AlphaCoV; 1st round	60,1°C-64,6°C
AlphaCoVG2 RGU R2 (10mM)	GAMCCYAAWAYCATSGCAGAGATCAT	Hemi-Nested G2 AlphaCoV; 2nd round	62,9°C-67,9°C
SturniraCoV RGU R1 (10mM)	CCACCCGTACAATGAACAACTTC	Hemi-Nested Sturnira CoV; 1st round	62,9°C
SturniraCoV RGU R2 (10mM)	CTTTGCACCYAGTATCATCGCTG	Hemi-Nested Sturnira CoV; 2nd round	62,9°C-64,6°C
BetaCoV RGU R1 (10mM)	TAGAAWCCACCATTDGARTAMACAAC	Hemi-Nested BetaCoV; 1st round	62,9°C-64,6°C
BetaCoV RGU R2 (10mM)	GAMCCYAAWAYCATSGCAGAGA CAT	Hemi-Nested BetaCoV; 1st round	62°C-62°C
SP3080 (10mM)	CTTCTTCTTTGCTCAGGATGGCAATGCTGC	Hemi-Nested BetaCoV; 1st and 2nd rounds	72,1°C
GrISP1 (10mM)	TTCTTTGCACAGAAGGGTGATGC	Hemi-Nested G1 and G2 AlphaCoV; 1st and 2nd rounds	62,9°C
GrISP2 (10mM)	CTTTGCACAAAAAGGTGATGCWGC	Hemi-Nested Sturnira CoV; 1st and 2nd rounds	63,6°C-63,6°C
Sequencing RGU sturnira R (10mM)	TCATATTAGGCAATGCACGG	Reverse primer specific for sequencing	56,4°C

A heminested RT-PCR was performed using two specific reverse primers, one designed in the beggining of the 394nt screening fragment achieved in the previous study of Góes et al., (2016) using the an adapted screening protocol published by Chu et al., (2011), and the second one in a region downstream of the same fragment (100nt). The consensus forward primer was previously designed by Drexler et. al. (2010). The PCR rounds for each group of samples consisted of the same quantities of each reagent, however, each assay had specific primers. The cycling protocols used for the first and second steps are shown in table 4 and 5.

Table 4 - Cycling protocol for the first round of the extension of the screening fragment.

Step	Temperature (°C)	Duration	Number of cycles
Reverse transcription	50°C	30 min	
Initial denaturation	94°C	2 min	
Denaturantion	94°C	15 sec	
Annealing	60°C*	15 sec	10X
Elongation	68°C	40 sec	
Denaturantion	95°C	15 sec	
Annealing	50°C	30 sec	40X
Elongation	72°C	40 sec	
Final Elongation	72°C	1 min	

* Touchdown of -1 °C per cycle. Temperature decreases 1°C per cycle.

Table 5 - Cycling protocol for the second round of the extension of the screening fragment.

Step	Temperature (°C)	Duration	Number of cycles
Initial denaturation	95°C	3 min	
Denaturation	94°C	15 sec	
Annealing	60°C*	20sec	10X
Elongation	72°C	40 sec	
Denaturation	95°C	15 sec	
Annealing	50°C	20sec	40X
Elongation	72°C	40 sec	
Final Elongation	72°C	1 min	

* Touchdown of -1 °C per cycle.

For the first round of the heminested PCR and for one reaction, 2,5 µL of the extracted RNA was added to a master mix constituted by the following reagents: 0,05 µL of RNase free water; 6,25 µL of 2X Reaction Mix; 0,2 µL of $MgCl_2$; 0,5 µL of BSA (1mg/mL); 0,2 µL of $MgSO_4$ (50mM); 0,5 µL of BSA (1mg/mL); 0,5 µL of Superscript III RT Mix and 1,25 µL of each forward and reverse primer assigned in the table 4.For the second round and for each reaction, 1 µL of first round PCR product was added to the following master mix: 17,65 of RNase free water; 2,5 µL of 10XPCR Buffer without $MgCl_2$; 0,5 µL of dNTP mix (10mM each); 0,75 µL of $MgCl_2$ (50 mM); 0,1 µL of Platinium *Taq* DNA Polymerase and 1,25 µL of each second-round primers assigned for each group assay. The second round PCR products were analyzed through gel electrophoresis as described in section 2.2 of the Materials and Methods.

2.5. DNA Sequencing

The samples with bands that correspond to the expected amplicon size were sent for sequencing to Microsynth SEQLAB Sequence Laboratories Goettingen GmbH. The DNA sequencing method was based on the dye terminator sequencing method (Sanger et al., 1977). The cleanup of the samples consisted on the addiction 5 µL of the second round PCR product to 2 µL of the reagent ExoSAP and incubation during five minutes at 37°C and finally inactivation by heat at 80°C during 10 minutes. 3 µL of the ExoSAP reaction was added to 5 µL of RNase free water and 2 µL of the second round reverse or forward primer. The results were analyzed in Geneious prime software version 2019.0.4 and compared to a public nucleotide database, using BLAST from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/).

2.6. Genome analysis

2.6.1. Genome assembling and phylogeny

A full genome from a betacoronavirus acquired from a *Pteronotus parnelli* bat present in the Institute and not published was characterized along with an alphacoronavirus from a *Phyllostomus discolor* bat acquired by Gustavo Góes during his stay at the Institute. The characterization included the prediction of the genome organization and their respective non-structural proteins (NSPs), phylogenetic analysis and species delineation to classify the viruses.

2.6.2. Prediction of the genome organization and phylogenetic analysis

For the prediction of the presumed ORFs, each genome was aligned using the MAFFT alignment (Katoh et al., 2002; Katoh & Standley, 2013). For the betacoronavirus alignment, the ICTV reference species of betacoronavirus was used and for the alphacoronavirus in addition to the ICTV reference sequences used, all alphacoronavirus sequences related to the hCoV and bat-related NL63 and 229E and the Wencheng Sm shrew CoV were used. The ORFs were predicted using the tool “Find ORF” in Geneious Prime software.

The predicted main ORFs were annotated and extracted along with the ORFs from the other sequences. The nucleotide sequences were translated into an aminoacid alignment. Amino acid Bayesian phylogenetic analysis for each ORF was calculated by Mr Bayes v3.1 software (Ronquist & Huelsenbeck, 2003), using WAG amino acid substitution model for amino acid Bayesian trees and 2,000,000 generations sampled every 100 steps. The amino acid trees were annotated in the TreeAnnotator v1.5 program, using a burn-in of 5000 and a posterior probability limit of 0.5. The annotated trees were visualized with FigTree v1.4 from the BEAST package (Drummond et al., 2012). The amino acid sequences were furthermore compared by BLAST comparison (Blastp) program for the analyze of the homologies. In addition, the unknown ORFs were predicted using the online software Phyre (Kelley et al., 2015).

The prediction of the TRS (Transcription Regulatory Sequences) start and end codons of each ORF and slippery sequences were done by comparison with the reference sequences.

2.6.3. Prediction of nonstructural proteins (NSPs)

For the NSPs prediction of the new genomes, the OFR1ab of both genomes were aligned with the references species sequences and the NSPs were annotated based on the amino acid comparison of the first and final amino acid residues and their position. Each NSP was extracted and translated into amino acids and realigned using MAFFT alignment. Each NSP and the prediction of the functional putative domains were compared using the Blastp feature for confirmation of the new NSPs.

2.6.4. Genome similarity plots, species delineation and detection of recombination events

According to the ICTV, in order to assign a new species or a new genus the following criteria are used: a) A rooted phylogeny and calculation of pair-wise distances for the conserved domains of the replicase proteins pp1ab: ADRP, NSP5 (3CLpro), NSP12 (RdRp), NSP13 (Hel), NSP14 (ExoN), NSP15 (NendoU) and NSP16 (O-MT); b) The new member can be considered a representative of a new genus when it does not cluster with any of the current genera and share less the 46% amino acid sequence identity in the replicase domains mentioned above, with any other member of the family. C) New members that share more than 90% amino acid sequence identity of the replicase domains are considered to belong to the same species. Therefore, similarity plots were generated using the full genome of the new viruses and compared with the reference sequences on the SSE software version 1.3 (Simmonds, 2012) using a sliding window of 900 and a step size of 800 nucleotides.

The pairwise identities were calculated using MEGA 6.0 (Tamura et al., 2013) using the domains of the predicted NSPs and also the domains that were previously concatenated using Geneious Prime software. For the detection of recombination events, the nucleotide alignment of the new viruses and the reference sequences was uploaded to the RDP4 software (Martin et al., 2015) and analyzed.

3. Results

3.1. Coronavirus detection in Neotropical bats

3.1.1. Screening for coronavirus

A total of 978 bat samples were collected by Gustavo Góes in different regions of Brazil: Amazonas (n=3), Paraná (n=190), São Paulo (n= 789) and Tocantins (n=1). Figure 7 and Table 6 represents the location and number of samples collected in each area and a list of samples used for this study.

Table 6 - List of samples used in the project

Family	Genus	No. Of Samples		Location
		Intestine	Swabs, serum	
Molossidae	*Cynomops*	11		São Paulo
	Eumops	116		São Paulo
	Molossops	3		São Paulo
	Molossus	383		São Paulo
	Nyctinomops	5		São Paulo
	Promops	1		São Paulo
	Tadarida	30		São Paulo
Phyllostomidae	*Artibeus*	119		Paraná
		34	1	São Paulo
			2	Amazonas
	Carollia	45		Paraná
		3	1	São Paulo
	Desmodus	55		São Paulo
	Glossophaga	63	3	São Paulo
		1		Tocantins
	Minom		1	Amazonas
	Phyllostomus	3		São Paulo
	Platyrrhinus	5	1	São Paulo
	Pygoderma	1		São Paulo
	Sturnira	26		Paraná
		4	8	São Paulo
Vespertilionidae	*Eptesicus*	17		São Paulo
	Lasiurus	5		São Paulo
	Myotis	36		São Paulo
Unknown		23		

[a]The unknown samples correspond to intestines where the identification of the species and the location was unavailable.

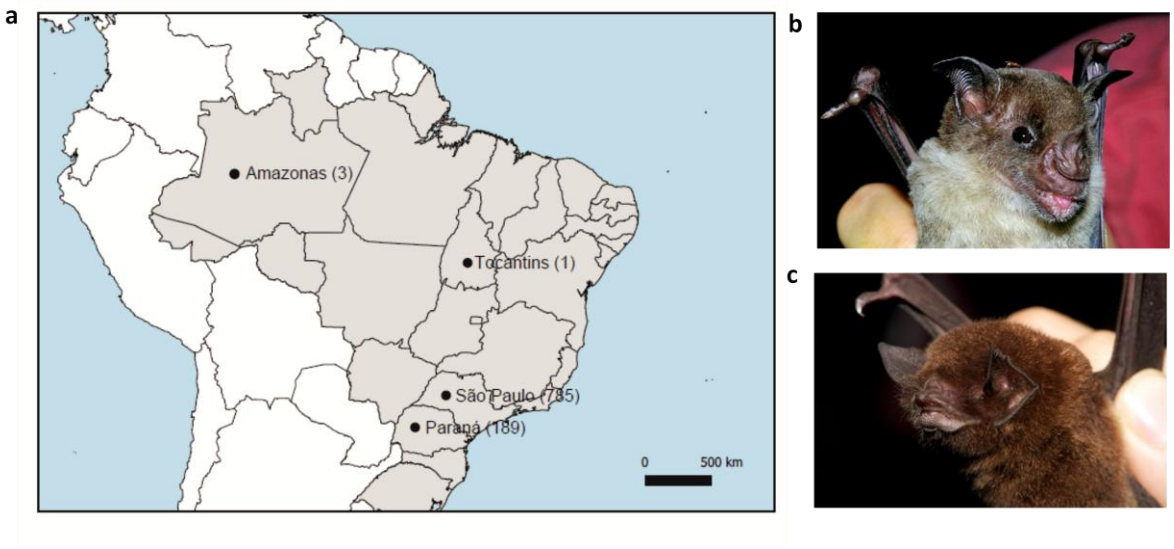


Figure 7 - Location and number of the samples collected as well as the bat species in which the full genomes where detected in the study. **(**a) Locations and number of the samples collected in Brazil. The number in parenthesis corresponds to the number of samples collected from bats. Amazonas n=3; Paraná n=189; São Paulo n=785 and Tocantins n= 1. (b) *Phyllostomus discolor* bat. Also known as Pale spear-nosed bat from the Phyllostomidae family is present in South and Central America. Access date and webpage: 15.07.2019 https://en.wikipedia.org/wiki/Pale_spear-nosed_bat. (c) *Pteronotus parnellii* bat. Also known as Parnell's mustached bat is a bat species that belongs to the Mormoopidae family is present in North, Central and South America. Access date and webpage: 15.07.2019 https://www.flickr.com/photos/23556954@N07/8709342848 (Simmons, 2005; Solari, 2016).

972 intestine samples were pooled into 122 pools of 8 and were screened for betacoronavirus. Out of the 122 pools, only 4 pools had a band with the corrected amplicon size after visualization in agarose gel. The PCR products were sent for sequencing and all were negative for betacoronavirus.

Recently many novel bat alphacoronavirus were detected. Regarding the alphacoronavirus screening assays, all published screening assays were analyzed in order to observe if some screening assay could cover all diversity of alphacoronavirus. An alignment was built with all ICTV reference sequences from the 4 CoV genera (alpha-, beta-, gamma- and deltacoronavirus) and with other important alphacoronavirus sequences for which complete genomes were available.

The primers from each screening assay were annotated and the primer binding regions of each screening assay and their mismatches to all reference alphacoronavirus were compared with the references. However, no screening assay was able to cover the alphacoronavirus' diversity due to the fact that the

primers used in the published screening assays only matched with some sequences and not the majority used in the allignment. Therefore, I started to develop a new assay but due to lack of time, the samples were not tested in this study.

3.1.2. Elongation of the screening fragment

In order to achieve the 816nt fragment from the RdRp, 35 already positive samples for CoV that were obtained by in Góes et al. (2016) were divided into 4 groups: Group I (G1) Alpha-CoV, Group II (G2) Alpha-CoV, Sturnira and Beta-CoV (Figure 8) and the RdRp fragment of 394bp was amplified in a heminested RT-PCR. For the Group 1 AlphaCoV (G1) (n=14), the 816nt fragment was obtained only from 11 samples. For the Group 2 AlphaCoV (G2) (n= 7), 5 samples had the correct amplicon size. Regarding the Sturnira Group (n=11), 3 RdRp fragments were amplified with the correct amplicon size and from the Beta Group (n=2) no amplification was achieved. Regarding the sequencing step for the samples from which the RdRp fragment was amplified (n=17), the sequencing results did not show results for CoV except for one sample from G1 AlphaCoV: *Phyllostomus discolor* intestine sample.

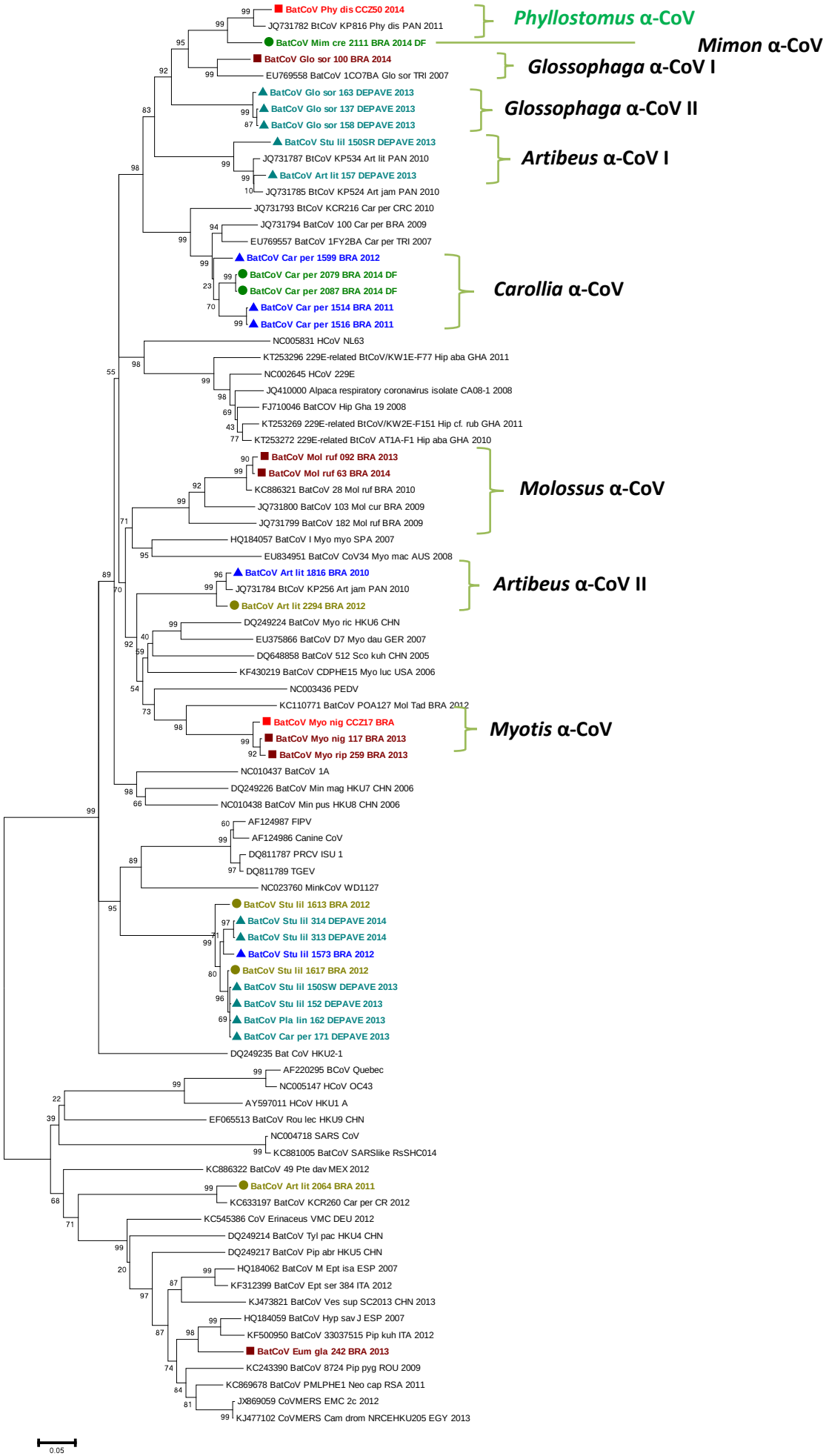

BatCoV Phy dis CCZ50 2014
JQ731782 BtCoV KP816 Phy dis PAN 2011
Phyllostomus α-CoV
BatCoV Mim cre 2111 BRA 2014 DF
Mimon α-CoV
BatCoV Glo sor 100 BRA 2014
EU769558 BatCoV 1CO7BA Glo sor TRI 2007
Glossophaga α-CoV I
BatCoV Glo sor 163 DEPAVE 2013
BatCoV Glo sor 137 DEPAVE 2013
BatCoV Glo sor 158 DEPAVE 2013
Glossophaga α-CoV II
BatCoV Stu lil 150SR DEPAVE 2013
JQ731787 BtCoV KP534 Art lit PAN 2010
BatCoV Art lit 157 DEPAVE 2013
JQ731785 BtCoV KP524 Art jam PAN 2010
Artibeus α-CoV I
JQ731793 BtCoV KCR216 Car per CRC 2010
JQ731794 BatCoV 100 Car per BRA 2009
EU769557 BatCoV 1FY2BA Car per TRI 2007
BatCoV Car per 1599 BRA 2012
BatCoV Car per 2079 BRA 2014 DF
BatCoV Car per 2087 BRA 2014 DF
BatCoV Car per 1514 BRA 2011
BatCoV Car per 1516 BRA 2011
Carollia α-CoV
NC005831 HCoV NL63
KT253296 229E-related BtCoV/KW1E-F77 Hip aba GHA 2011
NC002645 HCoV 229E
JQ410000 Alpaca respiratory coronavirus isolate CA08-1 2008
FJ710046 BatCOV Hip Gha 19 2008
KT253269 229E-related BtCoV/KW2E-F151 Hip cf. rub GHA 2011
KT253272 229E-related BtCoV AT1A-F1 Hip aba GHA 2010
BatCoV Mol ruf 092 BRA 2013
BatCoV Mol ruf 63 BRA 2014
KC886321 BatCoV 28 Mol ruf BRA 2010
JQ731800 BatCoV 103 Mol cur BRA 2009
JQ731799 BatCoV 182 Mol ruf BRA 2009
Molossus α-CoV
HQ184057 BatCoV I Myo myo SPA 2007
EU834951 BatCoV CoV34 Myo mac AUS 2008
BatCoV Art lit 1816 BRA 2010
JQ731784 BtCoV KP256 Art jam PAN 2010
BatCoV Art lit 2294 BRA 2012
Artibeus α-CoV II
DQ249224 BatCoV Myo ric HKU6 CHN
EU375866 BatCoV D7 Myo dau GER 2007
DQ648858 BatCoV 512 Sco kuh CHN 2005
KF430219 BatCoV CDPHE15 Myo luc USA 2006
NC003436 PEDV
KC110771 BatCoV POA127 Mol Tad BRA 2012
BatCoV Myo nig CCZ17 BRA
BatCoV Myo nig 117 BRA 2013
BatCoV Myo rip 259 BRA 2013
Myotis α-CoV
NC010437 BatCoV 1A
DQ249226 BatCoV Min mag HKU7 CHN 2006
NC010438 BatCoV Min pus HKU8 CHN 2006
AF124987 FIPV
AF124986 Canine CoV
DQ811787 PRCV ISU 1
DQ811789 TGEV
NC023760 MinkCoV WD1127
BatCoV Stu lil 1613 BRA 2012
BatCoV Stu lil 314 DEPAVE 2014
BatCoV Stu lil 313 DEPAVE 2014
BatCoV Stu lil 1573 BRA 2012
BatCoV Stu lil 1617 BRA 2012
BatCoV Stu lil 150SW DEPAVE 2013
BatCoV Stu lil 152 DEPAVE 2013
BatCoV Pla lin 162 DEPAVE 2013
BatCoV Car per 171 DEPAVE 2013
DQ249235 Bat CoV HKU2-1
AF220295 BCoV Quebec
NC005147 HCoV OC43
AY597011 HCoV HKU1 A
EF065513 BatCoV Rou lec HKU9 CHN
NC004718 SARS CoV
KC881005 BatCoV SARSlike RsSHC014
KC886322 BatCoV 49 Pte dav MEX 2012
BatCoV Art lit 2064 BRA 2011
KC633197 BatCoV KCR260 Car per CR 2012
KC545386 CoV Erinaceus VMC DEU 2012
DQ249214 BatCoV Tyl pac HKU4 CHN
DQ249217 BatCoV Pip abr HKU5 CHN
HQ184062 BatCoV M Ept isa ESP 2007
KF312399 BatCoV Ept ser 384 ITA 2012
KJ473821 BatCoV Ves sup SC2013 CHN 2013
HQ184059 BatCoV Hyp sav J ESP 2007
KF500950 BatCoV 33037515 Pip kuh ITA 2012
BatCoV Eum gla 242 BRA 2013
KC243390 BatCoV 8724 Pip pyg ROU 2009
KC869678 BatCoV PMLPHE1 Neo cap RSA 2011
JX869059 CoVMERS EMC 2c 2012
KJ477102 CoVMERS Cam drom NRCEHKU205 EGY 2013
0.05

Figure 8 - Phylogenetic tree of the 394bp of the RdRp of positive samples. The positive samples are coloured according to their groups. The *Phyllostomus discolor* partial sequence from de RdRp is represented in green. Group I AlphaCoV (G1) (n=14), Group II AlphaCoV (G2) (n= 7), Sturnira Group (n=11), Beta Group (n=2). This tree was kindly supplied by Gustavo Goés.

3.2. Genome annotation

For this study, 2 full genomes were annotated and analyzed: A betacoronavirus unpublished previously discovered in the bat species *Pteronotus parnellii* (Pte-BetaCoV) (Figure 9) and another one, corresponding to an alphacoronavirus found in *Phyllostomus discolor* (Phyl-AlphaCoV) the sequence of which was completed in the course of this study. Phyl-AlphaCoV was obtained by Gustavo Goés using Illumina sequencing method whereas the Pte-BetaCoV kindly made available by Andrea Rasche, was acquired by Illumina HiSeq2500.

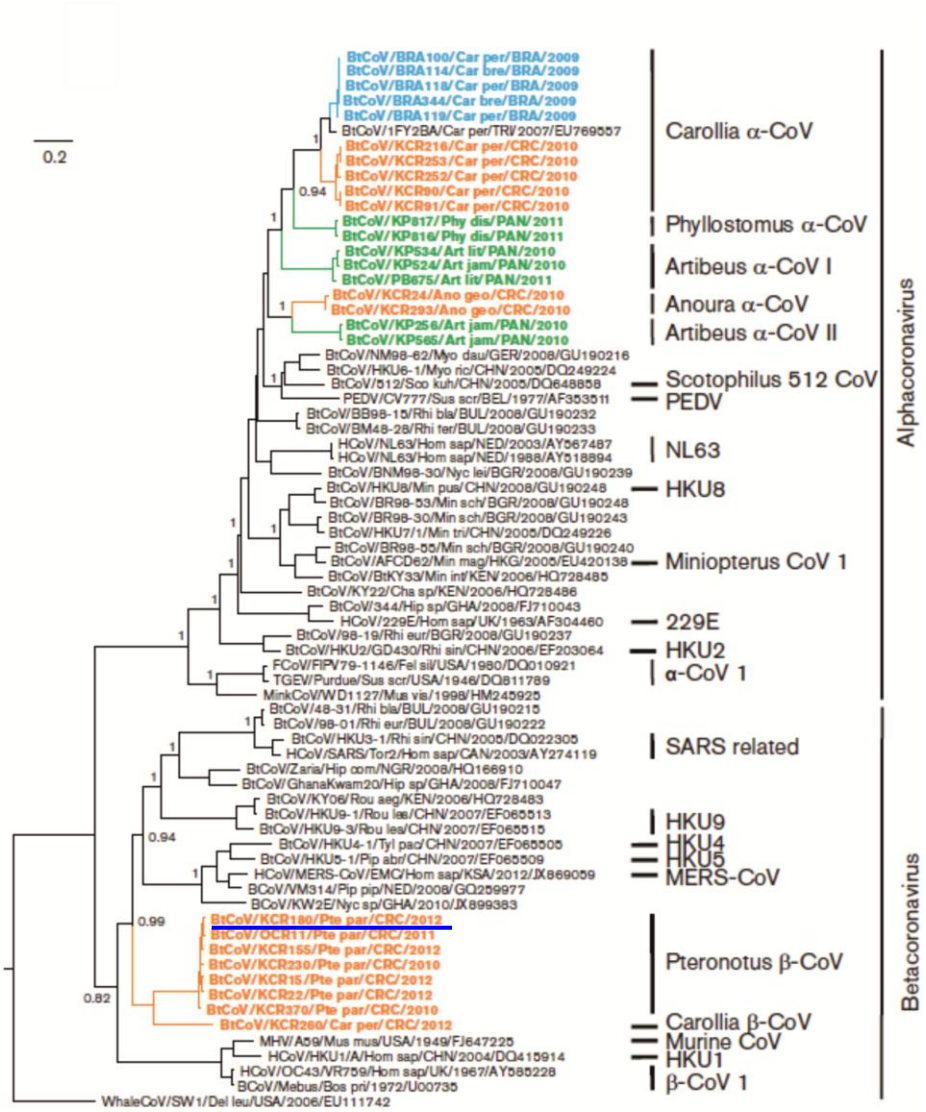


Figure 9 - Bayesian tree of the 816nt gap free of the translated RdRp fragment sequences. Pte-BetaCoV is represented in a blue underline (Corman et al., 2013).

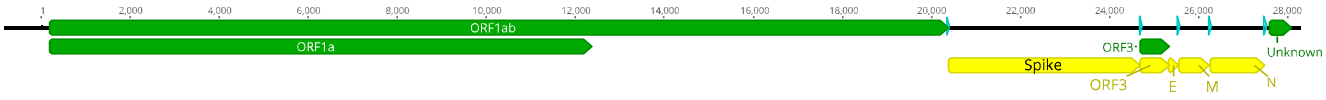
3.2.1. ORF prediction

The Pte-BetaCoV genome was aligned with the ICTV reference sequences and the Phyl-AlphaCoV was aligned with the ICTV reference sequences and with all alphacoronavirus related to the hCoV NL63 and 229E, related bat CoV 229E and Nl63 and Wencheng Sm shrew CoV. The ORFs were predicted using the Geneious prime software and by comparison with the found ORF with the annotated ORFs from the reference sequences.

For the Phyl-AlphaCoV genome, 7 ORFs were predicted (Figure 10a): ORF1ab, Spike gene, ORF3, E, M, N and 1 unknown ORF. The ORFs were compared with the starting and end amino acid positions of the reference sequences and also with the presence of a TRS (Transcription regulatory sequence) upstream of the beginning of each ORF. The ORFs sequences were analyzed by BLAST to observer there homology with other sequences. For the Pte-BetaCoV genome (Figure 10b), the ORFs were predicted using the same methodology for the prediction of the Phyl-AlphaCoV ORFs. The ORF 1ab, Spike gene, ORF3, E, M, N and also 1 unknown ORF were predicted. The ORF3 of this virus presented similarities with an Eidolon bat coronavirus (accession number: ADX59467) (Tao et al., 2012).

The amino acid sequences of the unknown ORFs in both genomes were blasted (blastp) and predicted using the online software Phyre and no result consisting of the putative OFRs function appeared.

(a)



(b)

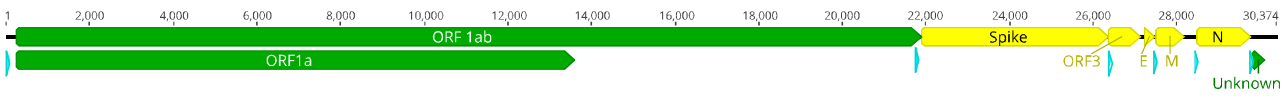


Figure 10 - Prediction of the genome organization of the complete genomes. In blue is represented the TRS. (a) Prediction of the organization of the Phyl-AlphaCoV genome. (b) Prediction of the organization of the Pte-BetaCoV genome.

3.2.2. TRS and ribosomal frameshift prediction

The Transcription Regulatory Sequence (TRS) was annotated based on the comparison of the TRS of the reference sequences. All the TRS found in both genomes were located upstream of each ORF. Table 7 and 8 contains the position and sequence of the putative TRS of both genomes. The leader TRS was only predicted for the Pte-BetaCoV and the TRS-B was found in 6 locations upstream of the *Pteronotus* ORFs and in 5 locations upstream of the Phyl-AlphaCoV ORFs.

Table 7 - Putative Transcription Regulatory Sequences (TRS) of the Phyl-AlphaCoV.

ORF	Nucleotide positions (start-end)	Number of amino acids	TRS Sequence
ORF 1*ab*	179-20,376	6732	No data
Spike	20,383-24,681	1433	TCTC**AACTAAG**TGAAA(9) AUG
ORF*3*	24,684-25,340	219	CGTC**AACTAAAC**ATGT(0)AUG
E	25,324-25,545	74	
M	25,552-26,244	231	ACGT**CTAAAC**GAAGA(4)AUG
N	26,261-27,493	411	AATC**AACTAAAA**ACA(6) AUG
Unknown 1	27,496-27,984	161	TCTC**AACTAAAA**ATGC (1) AUG

[a] Number in brackets represents the number of nucleotides to the putative start codon.

[b] The blank spaces in the table represent the absence of a putative TRS for the ORF in question. In ORF1ab no TRS was found due to the missing of part of the initial genome sequence.

[c] The TRS is represented in bold.

Table 8 - Putative Transcription Regulatory Sequences (TRS) of the Pte-BetaCoV.

ORF	Nucleotide positions (start-end)	Number of amino acids	TRS Sequence
ORF 1*ab*	247-21,905	7221	CCCG**AACTAACGAACT**AAA (210) AUG
Spike	21,893-26,368	1492	TTCC**AACTAAAACCAA**AGA (145) AUG
ORF*3*	26,376-27,140	255	TTAT**AACTAACAACCT**AAT (1) AUG
E	27,229-27,480	84	
M	27,497-28,180	228	TTGG**GTCTAACGAACT**TAA (10) AUG
N	28,465-29,775	437	TTAT**AACTAACTTACA**TCA (5) AUG
Unknown 2	29,790-30,104	105	CGTAA**ACAAACAAACA**TCA (3) AUG

[a] Number in brackets represents the number of nucleotides to the putative start codon.

[b] The blank spaces in the table represent the absence of a putative TRS for the ORF in question. In ORF1ab no TRS was found due to the missing of part of the initial genome sequence.

[c] The TRS is represented in bold.

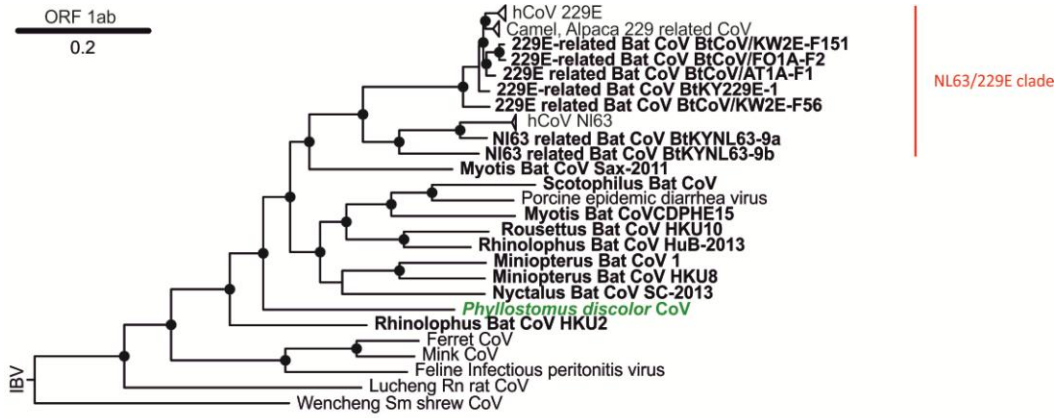
In both genomes, the ribosomal frameshift sequence present in the ORF1ab was predicted based on the slippery sequence found in the literature and in the reference sequences. The ribosomal frameshift sequence was annotated by the comparison of the amino acid frame of each ORF and the position of the stop codons in the reference and annotated sequence.

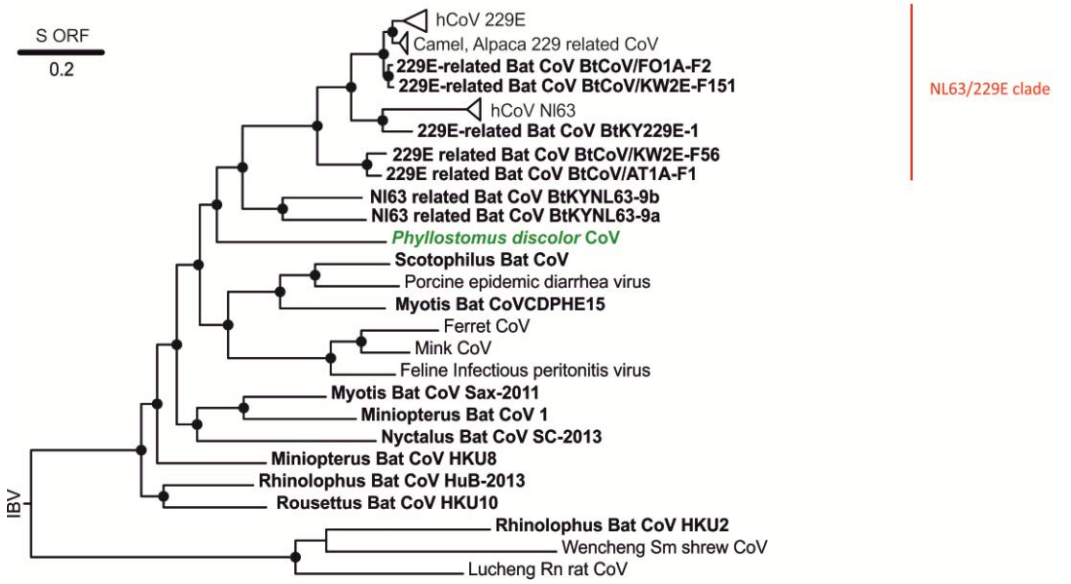
In the Phyl-AlphaCoV genome, the slippery sequence 5'-UUUAAAC-3' is located at the nucleotide positions 12,322 to 12,328 and in the Pte-BetaCoV genome, the slippery sequence with the same nucleotides was located in the nucleotide positions 13,570 to 13,575.

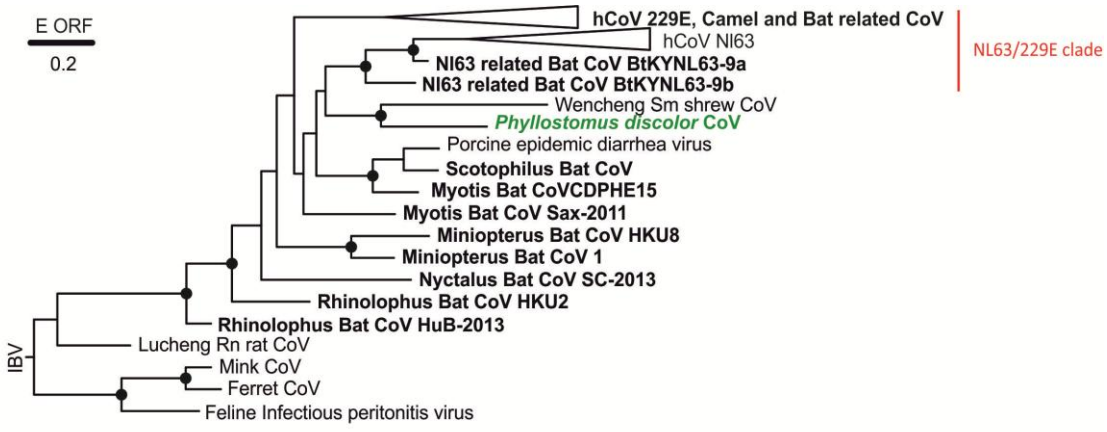
3.2.3. ORF phylogenetic analysis

Bayesian analysis of the ORFs 1ab, S, E, M and N were performed including all ICTV reference sequences. For the alphacoronavirus alignment, all viruses related to the hCoV NL63 and 229E and Wencheng Sm shrew CoV were additionally included. The addition of these viruses was due to the fact that preliminary phylogenetic analysis of the S ORF demonstrated that the Spike protein of Phyl-AlphaCoV clustered with NL63 and 229E related viruses. Therefore, all 229E and NL63 viruses were added. The Wencheng Sm shrew virus was also added to complement the diversity of alphacoronaviruses due to the fact that this virus is a divergent virus within this genus.

For the Phyl-AlphaCoV analysis, the gammacoronavirus Infectious bronchitis virus (IBV) was used as the outgroup for all the ORF analysis. The IBV was used instead of a betacoronavirus due to the fact that the S ORF of Rhinolophus bat CoV HKU2 presents similarities with SARS-related CoVs due to recombination events (Lau et al., 2007). The amino acid trees are presented in Figure 11. The amino acid trees of the ORFs 1ab and S showed supported basal nodes. However the amino acid trees of the ORFs E, M and N were not supported. In the ORF1ab phylogeny, the ORF1ab from Phyl-AlphaCoV clusters distantly with the other sister's clades of this group. However, The ORF S phylogeny shows that the Spike protein clusters with the group of the hCoV NL63 and 229E along with their respective related bat CoV.


ORF 1ab
0.2
hCoV 229E
Camel, Alpaca 229 related CoV
229E-related Bat CoV BtCoV/KW2E-F151
229E-related Bat CoV BtCoV/FO1A-F2
229E related Bat CoV BtCoV/AT1A-F1
229E-related Bat CoV BtKY229E-1
229E related Bat CoV BtCoV/KW2E-F56
hCoV NI63
NI63 related Bat CoV BtKYNL63-9a
NI63 related Bat CoV BtKYNL63-9b
Myotis Bat CoV Sax-2011
Scotophilus Bat CoV
Porcine epidemic diarrhea virus
Myotis Bat CoVCDPHE15
Rousettus Bat CoV HKU10
Rhinolophus Bat CoV HuB-2013
Miniopterus Bat CoV 1
Miniopterus Bat CoV HKU8
Nyctalus Bat CoV SC-2013
Phyllostomus discolor CoV
Rhinolophus Bat CoV HKU2
Ferret CoV
Mink CoV
Feline Infectious peritonitis virus
Lucheng Rn rat CoV
Wencheng Sm shrew CoV
IBV
NL63/229E clade


S ORF
0.2
hCoV 229E
Camel, Alpaca 229 related CoV
229E-related Bat CoV BtCoV/FO1A-F2
229E-related Bat CoV BtCoV/KW2E-F151
hCoV NI63
229E-related Bat CoV BtKY229E-1
229E related Bat CoV BtCoV/KW2E-F56
229E related Bat CoV BtCoV/AT1A-F1
NI63 related Bat CoV BtKYNL63-9b
NI63 related Bat CoV BtKYNL63-9a
Phyllostomus discolor CoV
Scotophilus Bat CoV
Porcine epidemic diarrhea virus
Myotis Bat CoVCDPHE15
Ferret CoV
Mink CoV
Feline Infectious peritonitis virus
Myotis Bat CoV Sax-2011
Miniopterus Bat CoV 1
Nyctalus Bat CoV SC-2013
Miniopterus Bat CoV HKU8
Rhinolophus Bat CoV HuB-2013
Rousettus Bat CoV HKU10
Rhinolophus Bat CoV HKU2
Wencheng Sm shrew CoV
Lucheng Rn rat CoV
IBV
NL63/229E clade


E ORF
0.2
hCoV 229E, Camel and Bat related CoV
hCoV NI63
NI63 related Bat CoV BtKYNL63-9a
NI63 related Bat CoV BtKYNL63-9b
Wencheng Sm shrew CoV
Phyllostomus discolor CoV
Porcine epidemic diarrhea virus
Scotophilus Bat CoV
Myotis Bat CoVCDPHE15
Myotis Bat CoV Sax-2011
Miniopterus Bat CoV HKU8
Miniopterus Bat CoV 1
Nyctalus Bat CoV SC-2013
Rhinolophus Bat CoV HKU2
Rhinolophus Bat CoV HuB-2013
Lucheng Rn rat CoV
Mink CoV
Ferret CoV
Feline Infectious peritonitis virus
IBV
NL63/229E clade

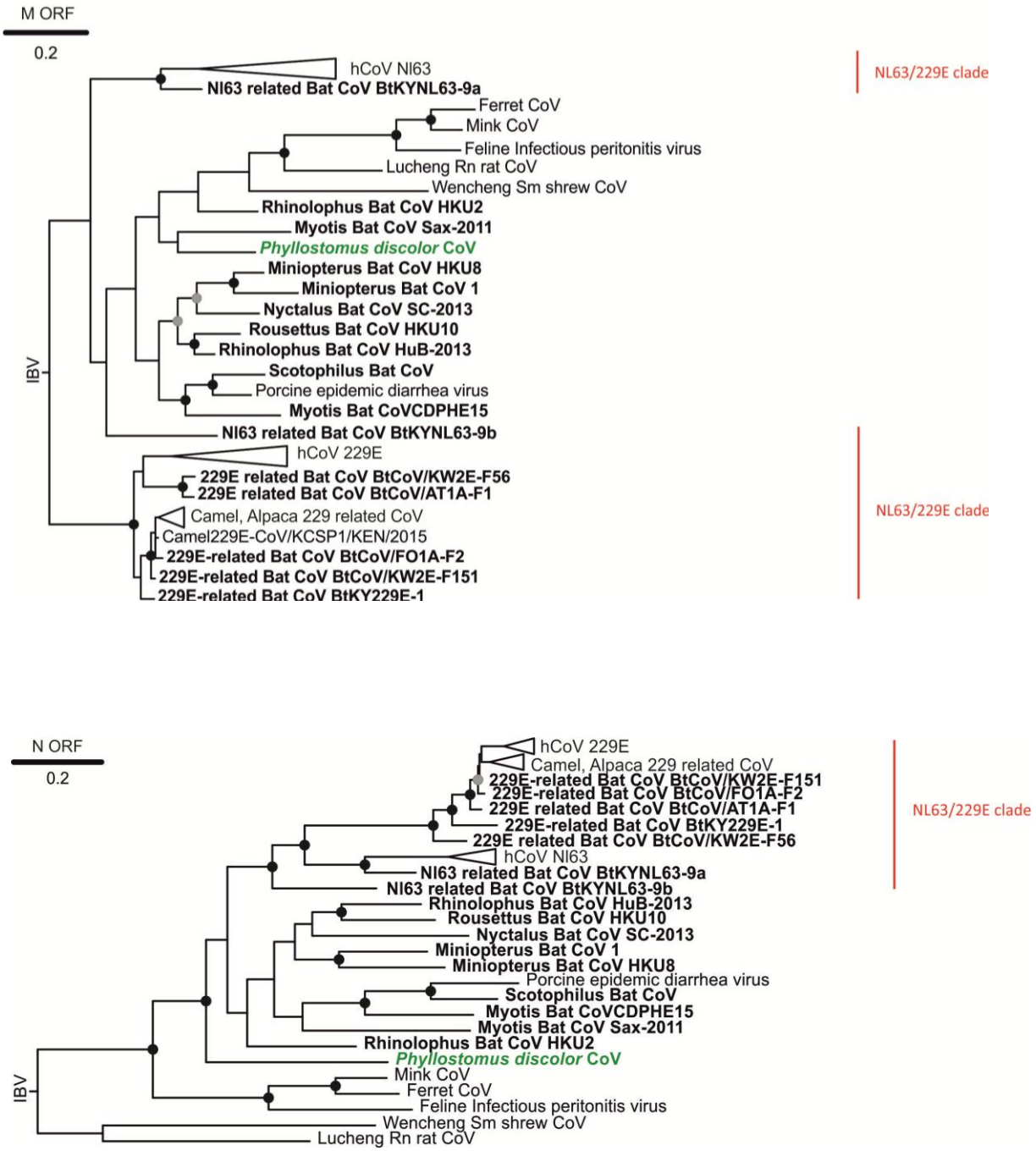
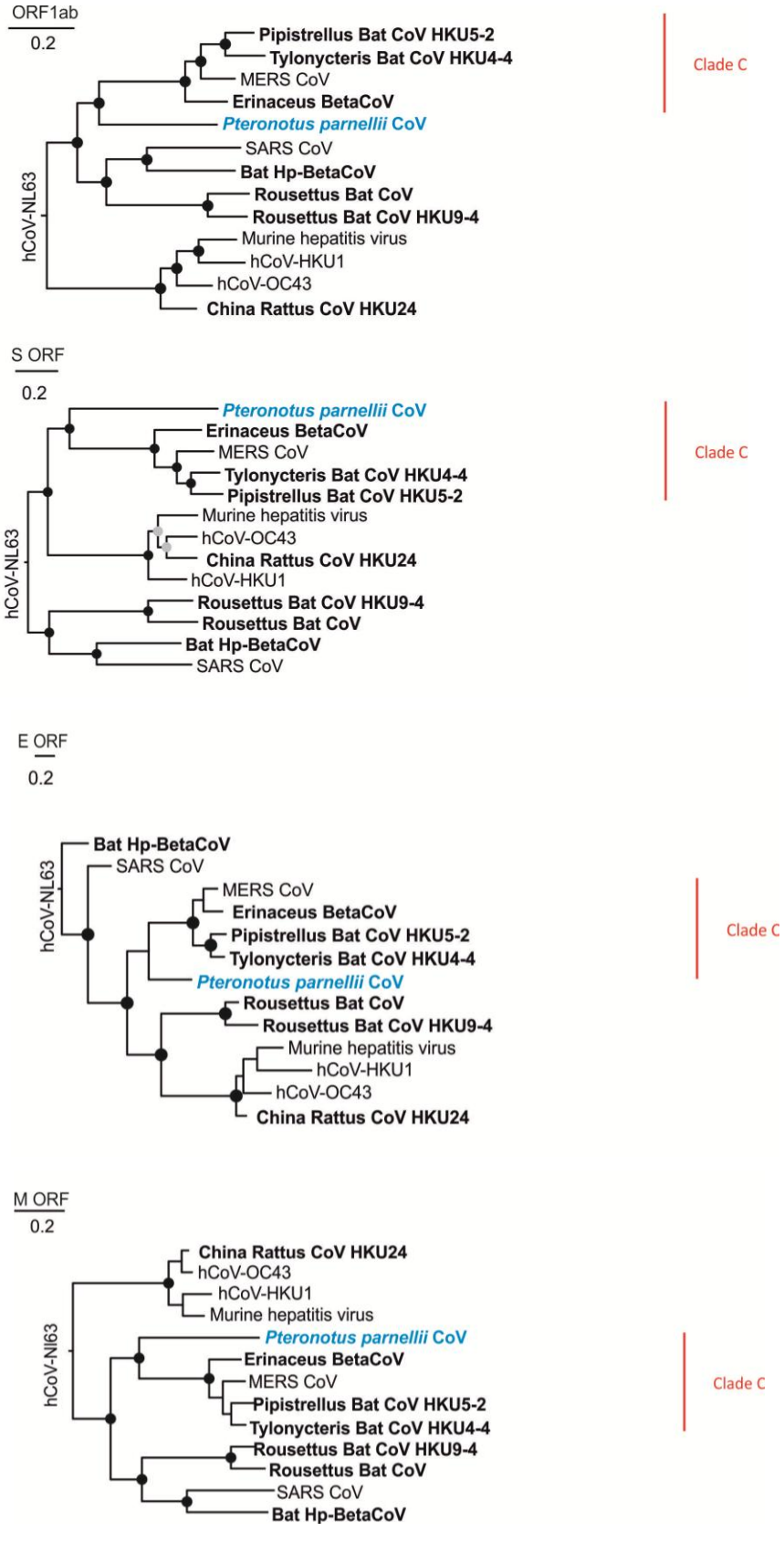


Figure 11 - Bayesian phylogenies of the major ORFs of the Phyl-AlphaCoV (Green). All bat CoV are shown in bold. The gammacoronavirus Infectious bronchitis virus (IBV) was used for rooting the trees. The collapsed branches include all complete genomes related to the clade NL63/229E. All complete genomes and their accession number are displayed in Table S1 in the Supplementary section. The Bayesian trees were calculated using a WAG amino acid substitution model. Values with a posterior probability above 0.95 are shown in black dots and values between 0.90 and 0.95 are shown in grey dots.

For the Pte-BetaCoV, the human alphacoronavirus NL63 was used to root the Bayesian phylogenetic trees as shown in Figure 12. OFRs E and M were statistically supported (not all nodes present a posterior probability superior to 0.90) whereas the other major ORFs were fully supported and the Pte-BetaCoV clustered distantly with clade C betacoronavirus group where the hCoV MERS, Erinaceus BetaCoV, HKU5 and HKU4 belong.



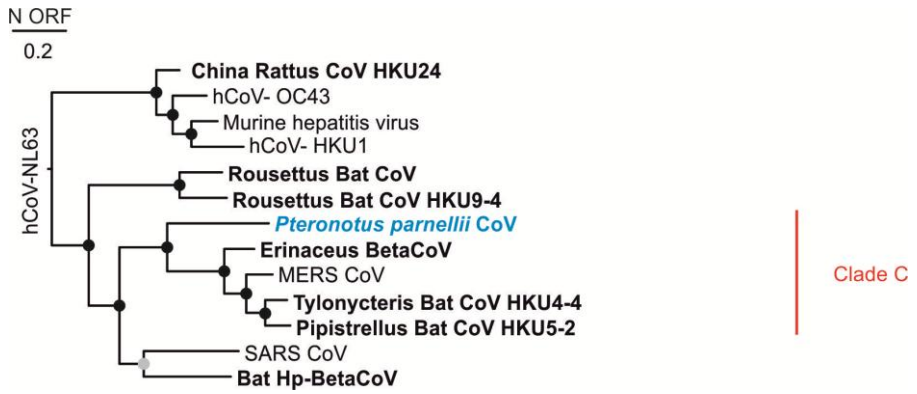


Figure 12 - Bayesian phylogenies of the ORFs of the Pte-BetaCoV (Blue). All bat CoV are shown in bold. The human alphacoronavirus NL63 was used for rooting the trees. The Bayesian trees were calculated using a WAG amino acid substitution model. Values with a posterior probability above 0.95 are shown in black dots and values between 0.90 and 0.95 are shown in grey dots.

3.2.4. Prediction of nonstructural proteins (NSPs)

To further delineate the characterization of the Phyl-AlphaCoV and Pte-BetaCoV genomes, the prediction of the putative polyprotein pp1a/1ab cleavage sites and their non-structural proteins (NSPs) was performed. Both genomes were aligned with the correspondent reference sequences. The locations of the 16 NSPs was done by comparison to the locations of the reference NSPs sequences and by the locations of other NSPs of genomes characterized in the literature. In addition, the N- and C-terminal amino acid residues of each new coding sequence were compared to the N- and C-terminal amino acid residues of the NSPs of the reference sequences. All NSPs from both genomes present a similar amino acid length to the reference NSPs however the nucleotide position of the NSPs- coding sequences differs in some cases. The prediction of the putative functional domains of the NSPs was done by associating the putative NSPs to the main function of the NSPs of the reference sequences and the function described in the literature. Table 9 and 10 provides the information about the protein length, location of the first and end amino acid residues and their putative functional domains.

Table 9 - Prediction of the putative pp1a/pp1ab cleavage sites of Phyl-AlphaCoV based on the comparison with the ICTV reference sequences. The putative function of each domain is based on the ICTV.

NSP	1st amino acid residue-last amino acid residue	Protein size (no. of amino acids)	Putative functional domain(s)
NSP 1	Met[1]-Gly[110]	110	IFN antagonist; Degradation of host mRNA; Inhibition of translation; Cell cycle arrest
NSP 2	Asn[111]- Gly[777]	667	
NSP 3	Gly[778]-Gly[2449]	1,672	ADRP;PL1[pro]; PL2[pro]
NSP 4	Ala[2450]-Gln[2934]	485	
NSP 5	**Ser[2935]-Gln[3237]**	**303**	**3CLpro**
NSP 6	Cys[3238]-Gln[3515]	278	
NSP 7	Ser[3516]-Gln[3598]	83	ssRNA binding
NSP 8	Ser[3599]-Gln[3793]	195	Noncanonical "secondary" RdRp with putative primase activity; forms hexadecameric supercomplex with NSP7
NSP 9	Asn[3794]-Gln[3902]	109	ssRNA binding; associates with RTCs
NSP 10	Ala[3903]-Met[4036]	134	Dodecameric zinc finger protein; associates with RTCs, simulates NSP 16 methyltransferase activity
NSP 11	Ser[4038]-Asp[4056]	19	Short peptide at the end of ORF*1a*
NSP 12	**Gln[4037]-Gln[4966]**	**930**	**RdRp**
NSP 13	**Ser[4967]-Gln[5563]**	**597**	**Hel, NTPase**
NSP 14	**Ala[5564]-Gln[6082]**	**519**	**ExoN,NMT**
NSP 15	**Gly[6083]-Gln[6429]**	**347**	**NendoU**
NSP 16	**Ser[6430]-Lys[6729]**	**300**	**OMT**

[a] Superscript number represents the positions of the nonstructural proteins in the polyprotein pp1a/pp1ab with the assumption of the ribosomal frameshift based on the slippery sequence (5'-UUUAAAC-3') present in CoV. The underlined NSPs indicate the NSPs used for the speciation criteria of the ICTV.

[b] IFN, Interferon; ADRP, ADP-ribose 1-phosphatase; PL1 [pro], papain-like protease 1; PL2 [pro], papain-like protease 2; 3CLpro, 3C-like main protease; RdRp, RNA-dependent RNA-polymerase; RTC, replicase/transcriptase complex; HEL, helicase; NTPase, nucleoside triphosphatase; ExoN, 3'-to-5' exoribonuclease; NMT, N7 methyltransferase; NendoU, nidoviral uridylate-specific endoribonuclease; OMT, ribose-2'-O-methyltransferase.

[c] The blank spaces represent the unknown function of the corresponded NSPs.

Table 10 - Prediction of the putative pp1a/pp1ab cleavage sites of Pte-BetaCoV based on the comparison with the ICTV reference sequences. The putative function of each domain is based on the ICTV.

NSP	1st amino acid residue-last amino acid residue	Protein size (no. of amino acids)	Putative functional domain(s)
NSP 1	Met[1]-Gly[209]	209	IFN antagonist; Degradation of host mRNA; Inhibition of translation; Cell cycle arrest
NSP 2	Val[210]- Gly[859]	650	
NSP 3	Ala860-Gly[2794]	1,935	ADRP;PL1[pro] ;PL2[pro]
NSP 4	Ala[2795]-Gln[3306]	512	
NSP 5	**Ala[3307]-Gln[3610]**	**304**	**3CLpro**
NSP 6	Gly[3611]-Gln[3901]	291	
NSP 7	Ser[3902]-Gln[3984]	83	ssRNA binding
NSP 8	Ser[3985]-Gln[4182]	198	Noncanonical "secondary" RdRp with putative primase activity; forms hexadecameric supercomplex with NSP7
NSP 9	Asn[4183]-Gln[4294]	112	ssRNA binding; associates with RTCs
NSP 10	Ala[4295]-Gln[4435]	141	Dodecameric zinc finger protein; associates with RTCs, simulates NSP 16 methyltransferase activity
NSP 11	Ala[4436]-Ser[4449]	14	Short peptide at the end of ORF*1a*
NSP 12	**Ala[4436]-Gln[4966]**	**930**	**RdRp**
NSP 13	**Ser[5366]-Glu[5959]**	**594**	**Hel, NTPase**
NSP 14	**Ala[5990]-Gln[6518]**	**529**	**ExoN,NMT**
NSP 15	**Ser[6519]-Gln[6915]**	**397**	**NendoU**
NSP 16	**Ala[6916]-Asn[7219]**	**304**	**OMT**

[a] Superscript number represents the positions of the nonstructural proteins in the polyprotein pp1a/pp1ab with the assumption of the ribosomal frameshift based on the slippery sequence (UUUAAAC) present in CoV. The underlined NSPs indicate the NSPs used for the speciation criteria of the ICTV.

[b] IFN, Interferon; ADRP, ADP-ribose 1-phosphatase; PL1 [pro], papain-like protease 1; PL2 [pro], papain-like protease 2; 3CLpro, 3C-like main protease; RdRp, RNA-dependent RNA-polymerase; RTC, replicase/transcriptase complex; HEL, helicase; NTPase, nucleoside triphosphatase; ExoN, 3'-to-5' exoribonuclease; NMT, N7 methyltransferase; NendoU, nidoviral uridylate-specific endoribonuclease; OMT, ribose-2'-O-methyltransferase.

[c] The blank spaces represent the unknown function of the corresponded NSPs.

3.2.5. ORF amino acid identities

The ICTV speciation criteria assign that less than 46% of amino acid identity within the ICTV domains, a new genus is assigned and more than 90% of an amino acid identity the sequence is similar. With these criteria, the amino acid sequences of the major ORFs and also the concatenated domains of the major ORFs were analyzed using the Mega 6.0 software. In tables 11 and 12 are represented the amino acid differences per codon between the new coding sequences and their corresponding reference sequences.

In Phyl-AlphaCoV case (Table 11) the amino acid difference ranged from 35% to 43% which corresponds to 65% to 57% of amino acid identity respectively. For the Pte-BetaCoV genome (Table 12) the amino acid difference ranged from 33% to 40% which corresponds to 67% to 60% of amino acid identity respectively. The amino acid identity of both coding sequences analyzed by the ICTV criteria demonstrated that the 2 novel genomes do not correspond to members of a new genus, but do correspond to new-species of previously defined genera. In the case of the Pte-BetaCoV, it is related to the clade C of betacoronavirus and in the case of the Phyl-AlphaCoV to the NL63/229E clade of the alphacoronavirus.

Table 11 - Comparison of amino acid differences of the Phyl-AlphaCoV genome.

	Replicase Polyprotein Domain							
Phyl-AlphaCoV vs:	**NSP3**	**NSP5**	**NSP12**	**NSP13**	**NSP14**	**NSP15**	**NSP16**	**Concatenated domains**
Rousettus Bat CoV HKU10	56%	35%	19%	**18%**	29%	30%	28%	**35%**
FIPV	65%	37%	**27%**	26%	31%	34%	**34%**	41%
Swine enteric CoV	64%	38%	26%	26%	32%	33%	**34%**	41%
TGEV	64%	37%	**27%**	26%	32%	33%	**34%**	41%
Mink CoV	60%	37%	26%	24%	32%	33%	31%	39%
Rhinolophus Bat CoV HKU2	58%	35%	23%	21%	28%	30%	**27%**	37%
Ferret CoV	61%	37%	26%	24%	31%	34%	31%	39%
Myotis Bat CoV CDPHE15	57%	38%	22%	22%	29%	26%	30%	37%
PEDV	57%	**34%**	21%	19%	29%	29%	28%	36%
Scotophilus Bat CoV	57%	36%	22%	21%	30%	27%	29%	37%
Miniopterus Bat CoV HKU8	**54%**	35%	20%	21%	**27%**	34%	29%	**35%**
Miniopterus Bat CoV 1	56%	35%	19%	21%	30%	30%	28%	36%
Rhinolopus Bat CoV HuB-2013	56%	38%	19%	**18%**	29%	30%	**27%**	**35%**
Nyctalus bat CoV SC-2013	57%	38%	**18%**	21%	28%	29%	30%	36%

hCoV 229E	58%	35%	22%	19%	29%	28%	28%	36%
hCoV NL63	59%	36%	22%	19%	29%	29%	**27%**	37%
NL63-related Bat CoV	58%	32%	19%	**18%**	29%	**25%**	29%	**35%**
Myotis Bat CoV Sax-2011	56%	36%	20%	21%	28%	27%	**27%**	**35%**
Lucheng Rn Rat CoV	**67%**	**44%**	25%	**28%**	**36%**	**40%**	32%	**43%**

[a] The number in bold represent the lowest and the highest percentage of amino acid differences per site.

[b]FIPV, Feline Infectious peritonitis virus; TGEV, Transmissible gastroenteritis virus; PEDV Porcine epidemic diarrhea virus.

Table 12 - Comparison of amino acid differences of the Pte-BetaCoV genome.

	Replicase Polyprotein Domain							
Pte-BetaCoV vs:	**NSP3**	**NSP5**	**NSP12**	**NSP13**	**NSP14**	**NSP15**	**NSP16**	**Concatenated domains**
Tylonycteris Bat CoV HKU4-4	**50%**	43%	**26%**	27%	**34%**	**43%**	31%	**33%**
Pipistrellus Bat CoV HKU5-2	53%	42%	**26%**	**26%**	**34%**	46%	33%	**33%**
MERS CoV	52%	41%	**26%**	**26%**	**34%**	48%	**29%**	**33%**
Erinaceus BetaCoV	55%	43%	**26%**	**26%**	35%	45%	31%	**33%**
Bat Hp-BetaCoV	57%	**39%**	27%	27%	37%	46%	35%	34%
SARS CoV	56%	46%	27%	27%	36%	49%	37%	35%
Rousettus Bat CoV GCCDC1	54%	**50%**	28%	27%	38%	**57%**	**41%**	37%
Rousettus bat CoV HKU9-4	53%	48%	27%	**26%**	37%	55%	**41%**	36%
China Rattus CoV HKU24	84%	48%	30%	30%	39%	51%	33%	38%
hCoV-OC43	**86%**	48%	**32%**	30%	39%	52%	34%	39%
Murine hepatitis virus	85%	47%	**32%**	**32%**	**40%**	54%	35%	**40%**
hCoV-HKU1	83%	47%	**32%**	31%	**40%**	53%	35%	**40%**

[a] The number in bold represent the lowest and the highest percentage of amino acid differences per site.

3.2.6. Full genome similarity plots and species delineation

To further characterize the genomes and to assess the differences in the genomes at the nucleotide level, pairwise identity similarity plots of the full nucleotide genome was generated. In figures 13 and 14 is represented the similarity plots of the Phyl-AlphaCoV and Pte-BetaCoV genomes. Both genomes are about equidistant to reference viruses along the complete genome.

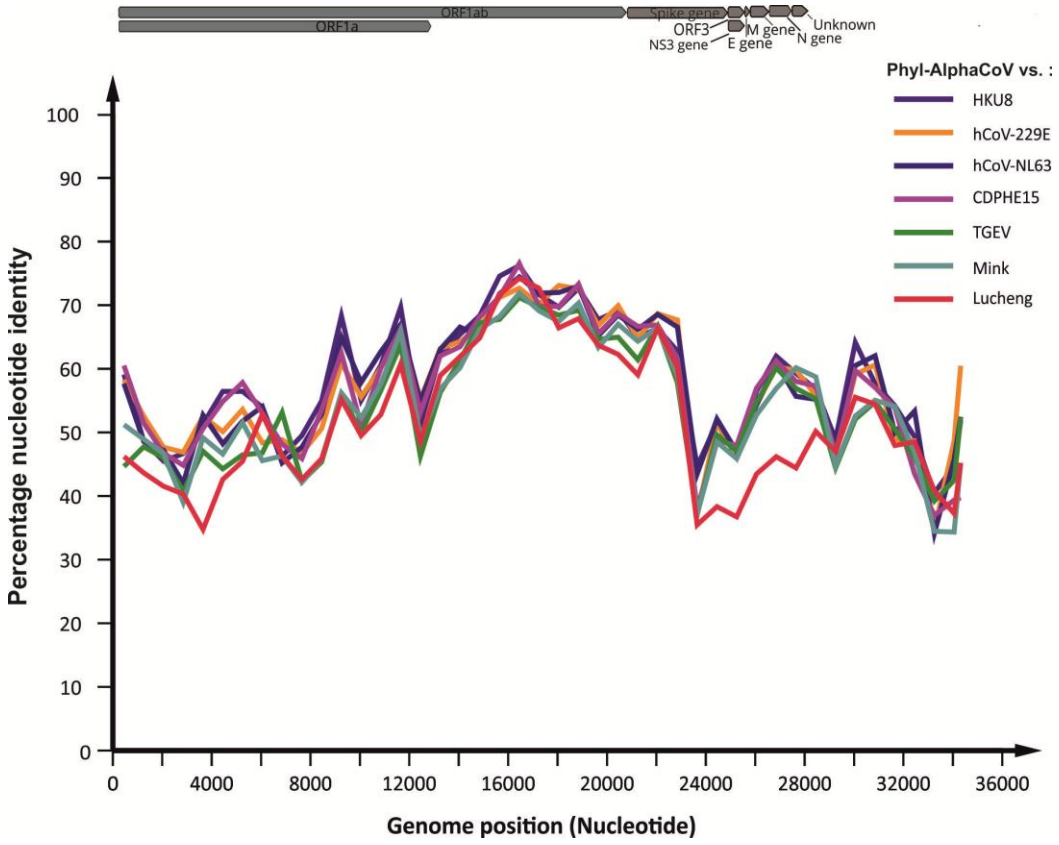


Figure 13 - Nucleotide sequence identity between Phyl-AlphaCoV and other selected alphacoronavirus sequences. Genome organization is displayed above the similarity graph. The plot was generated with a window size of 900 and a step size of 800 by using the SSE software version 1.3.

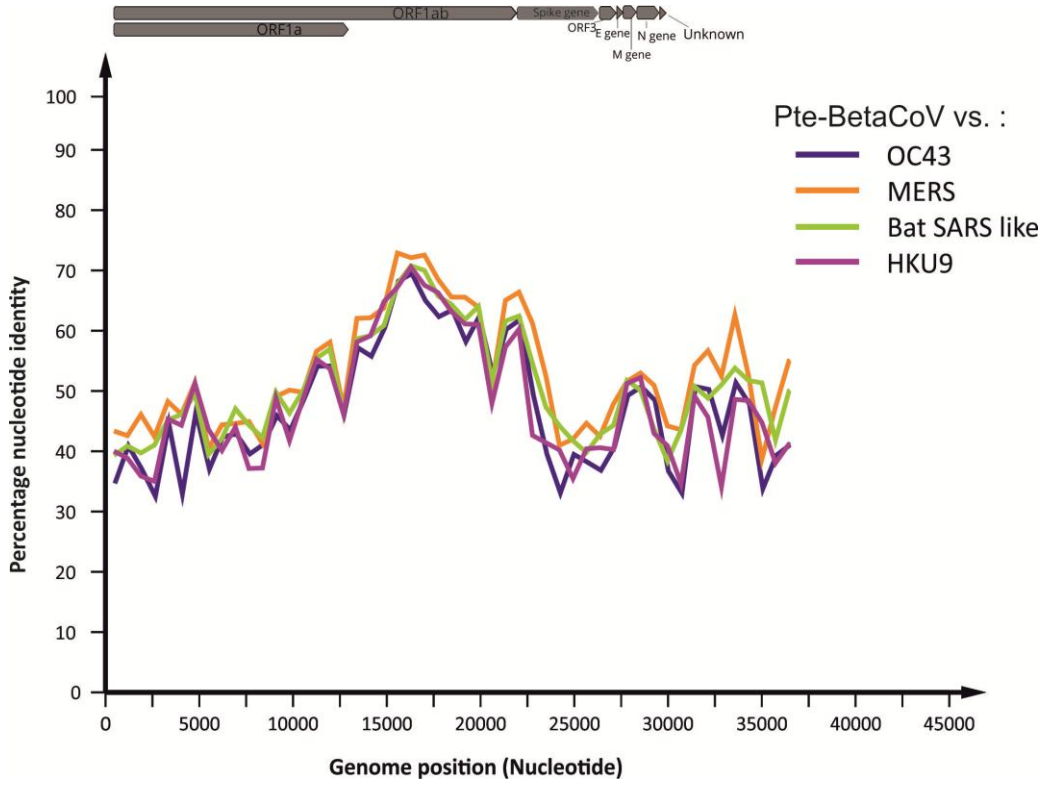


Figure 14 - Nucleotide sequence identity between Pte-BetaCoV and one sequence from each betacoronavirus clade. Genome organization is displayed above the similarity graph. The plot was generated with a window size of 900 and a step size of 800 by using the SSE software version 1.3.

The Phyl-AlphaCoV genomes showed a more conserved region that ranged from approximately 45% to 78% that corresponds to the ORF1b. The other regions were shown to be less conserve and corresponded to the other major ORFs.

The Pte-BetaCoV genome showed a similarity plot similar to the Phyl-AlphaCoV where the region more conserved ranged from approximately 42% to 71% corresponding to the ORF1b among all genomes. The other regions showed less conservation.

3.2.7. Detection of Recombination events

For the detection of recombination events, a nucleotide alignment of the full genome of the reference sequences and the corresponding full genome was analyzed on the RDP4 software with default settings. Evidence of recombination events was supported if at least 2 of the methods implemented in the package detected an event of recombination. Both genomes did not show clear evidence of recombination events.

4. Discussion and Conclusion

4.1. Detection of coronavirus in Neotropical bats

In this study, 972 samples from Brazilian bats were screened for CoV with the goal of assessing the diversity of CoVs in Neotropical bats. From this part of the study, an alphacoronavirus present in a *Phlyllostomus discolor* bat sample was provided for analysis by Gustavo Góes.

However, due to the lack of time, the analysis of the diversity of CoV was not completed due to the fact that these samples were only screened for one genus, betacoronavirus. In these samples there were a small number of positives for betacoronavirus in the screening PCR assay, however, the sequencing results showed negative results. This scenario might be due to a possible wrong amplification of the screening fragment which indicates an absence of betacoronavirus RNA in those samples. Regarding the alphacoronavirus screening assay, due to the high diversity of alphacoronavirus species, there is no suitable screening protocol that can detect efficiently the presence of all alphacoronavirus species. For that reason, it is important to develop a new screening assay which is adequate for the high diversity of this genus. Therefore the screening assay for alphacoronavirus was not performed in this thesis.

In the elongation of the screening fragment, samples previously screened for CoV were used to obtain the RdRp Grouping units (RGU) fragment using multiple combinations of primers defined in table 3. From all the samples used for this part of the project we were not able to elongate the screening fragment except for one sample, an alphacoronavirus detected in a *Phyllostomus discolor* bat from Brazil. For the rest of the samples, the heminested RT-PCR amplified a fragment with the correct amplicon size, however, all sequencing results were negative for coronavirus. In the chromatograms obtained from each positive sample, there were several problems that could be addressed to the reason why coronavirus was not detected. For instance, the presence of overlapping peaks might indicate possible the presence of a clone, contamination, excess of DNA or an unspecific amplification.

Another possible cause was the initial heminested PCR protocol, where the cycling temperatures were too low. The cycling temperature influences the primer binding and in this case, they could have bound to something unspecific and amplified the wrong fragment. Although several attempts were done with different primer combination and with different temperatures the results did not change. Also, an important point to observe is the number of Inosines in the degenerated primers and when the forward primers were designed. The degenerated primers enable the amplification of a region that does not possess a conserved sequence to design a primer. Therefore the more Inosines the primer has, the higher the probability of binding inaccurately to another fragment or to self complement.

The reverse primers used for these assays possess wobbles and that might have allowed the amplification of a wrong amplicon or self-complementation. Concerning the forward primers, these primers were designed in Drexler et al., 2010 and the diversity of CoV discovered afterwards is vast and therefore might not be able to bind efficiently to a larger diversity of CoV genomes.

4.2. Genome annotation

From this study 2 complete genomes were analyzed and annotated, one from a betacoronavirus present in a Costa Rican *Pterpnotus parnellii* bat (Pte-BetaCoV) and another from an alphacoronavirus present in a Brazilian *Phyllostomus discolor* bat (Phyl-AlphaCoV). Until now, no full genome was acquired from CoV present in New World bats.

The annotation of the genomes was successful and phylogenetic analysis demonstrated some ancient evolutionary relationships between these bat CoVs and their relatives. The ORF3 of Pte-BetaCoV showed some resemblances to an Eidolon bat coronavirus, a betacoronavirus detected in Kenya (Tao et al., 2012). Regarding these similarities further analysis must be conducted.

In both cases, ORFs were detected which were not present in any reference CoV. Hence, the function of these ORFs remains unclear. The existence of the ORFs was confirmed by the detection of TRS upstream of the ORFs. The presence of these unknown ORFs indicates a need to further analyze these genes to find more useful information to fully understand the alphacoronavirus genus.

4.2.1. Phylogenetic reconstructions

In terms of the phylogenetic analyses, some trees were not fully supported such as the Bayesian trees for the ORFs E, M and N in the Phyl-AlphaCoV case and for the ORFs E and M in the Pte-BetaCoV case. This might be explained by the small length of these ORFs and resulting short amino acid sequence information, which did not yield statistically supported phylogenies. Therefore no conclusions could be made about the relationship of the respective bat CoV and their relatives. However, for the other ORFs analyzed there are some extrapolations that can be made. In the case of the supported amino acid trees OFR1ab and Spike from Phyl-AphaCoV, the sequences cluster with different host groups within the alphacoronavirus genus.

In contrast, the ORF1ab is the most conserved ORF encoded by the genome and the Spike gene is the most diverse within the genome where the most frequent recombination events occur (Tao et al., 2017). Recombination events have been reported in emerging human betacoronavirus such as MERS, SARS and human alphacoronavirus NL63 and OC43, however, most reports describe these events between closely related viruses.

CoV posses a high frequency of homologous recombination mainly due to the discontinuous replication of RNA. The rapid change of their genetic constitution occurs, originating in some part to recombination events (Hon et al., 2008; Lau et al., 2011; Lau et al., 2010; Pyrc et al., 2006; Sabir et al., 2016). Therefore a recombination event in the *Spike* gene might possibly be the reason why these 2 sequences from the ORFs cluster with different groups. Regarding the ORF S phylogeny, this tree showed that this bat alphacoronavirus clusters with the hCoVs NL63, 229E and their relative bat CoVs. With these results, might be possible to infer an ancient relationship between this bat and the group of Nl63 and 229E like viruses which is supported by the presence of supported long branches in the phylogeny which is an indicator of a strong and long history between Phyl-AlphaCoV and the NL63 and 229E related viruses. Possibly this recombination event between Phyl-AlphaCoV and NL63 and 229E related CoVs occurred prior to the emergence of these viruses. The recombination analysis performed did not show any signs of recombination between these viruses. This result might be at odds with this explanation so further analysis must be done in order to prove this explanation.

In Corman et al., 2013 the Pte-BetaCoV did not cluster with the clade C betacoronavirus. This new data weakens the RGU phylogeny previously done with the 816nt RdRp fragments where the Pte-BetaCoV appeared to the divergent, or in other words a new clade.

For the Pte-BetaCoV, the supported amino acid trees of the ORFs 1ab, *Spike* gene and N reveal a distant sister relation between this virus and the group C betacoronavirus which includes MERS-CoV species. Although the most recent common ancestor of MERS-CoV still remains unknown, until now the most distant relative to MERS-CoV was found in hedgehog *Erinaceus* genus, within the clade C (Corman et al., 2014). This ancient sister relation in terms of evolution with the bat *Pteronotus parnellii* might be evidence for an ancestral association between MERS-CoV and bats (Wong et al., 2019).

MERS-CoV has been found only in bat families from the Old World such as the family Vespertilionidae, Molossidae and in the family Emballonuridae which occurs to the Old and New World (Drexler et al., 2014; Memish et al., 2013). The *Pteronotus parnellii* bat species belong to the family Mormoopidae which only occurs in the New World.

Bats only can migrate across short distance so therefore could never cross oceans except when they are retained in a boat. The passage of viruses from the Old World bats to New World bats might have been enabled by globalization by boats and planes (Constantine, 2003). With this fact is possible to establish a link connecting the clade C to the New World bats. The Pte-BetaCoV hints at a common ancestry of this virus and the clade C. However, virus exchange via boats and other vessels between New and Old World can not be discarded. This would allow a much later origin of this virus clade. This might also give a hint in the origins of this clade, which gives more support the bat origin of the clade C since the previous most divergent virus belonging to the clade originates from the insectivorous Eulipotyphla order namely hedgehog.

The hypothesis of the Pte-BetaCoV being and ancestor clade of the clade C is also supported by the amino acid identity that ranged from 60% to 67% which according to the ICTV speciation criteria this bat CoV does not form a new genus, however, it also does not belong to the same species.

In summary, the main goal of this thesis was to study the diversity of CoV in the New World and it gave access to important information. For instance, the development of a suitable new screening assay for alphacoronavirus that is capable to cover all diversity is a major point. Also, this project revealed important findings of the origins of important emerging CoV like the hCoV Nl63, 229E and MERS-CoV and their evolution with their bat hosts. The Phyl-AlphaCoV can possibly have an ancient relationship to the human and bat-related CoVs Nl63 and 229E due to a recombination event between these viruses and their phylogeny with Spike protein. However, in order to provide further arguments with the presence and relationship of ORF8 between Phly-AlphaCoV and these hCoVs further information is required. The ORF8 is an important marker to determine their relationship because is a conserved ORF in the hCoV 229E and Alpaca 229E related-CoV (Corman et al., 2015). The Pte-BetaCoV is also a hint to the origins of the clade C betacoronavirus due to their phylogenies observed in the main ORFs.

5. References

Anindita, P. D., Sasaki, M., Setiyono, A., Handharyani, E., Orba, Y., Kobayashi, S., Kimura, T. (2015). Detection of coronavirus genomes in Moluccan naked-backed fruit bats in Indonesia. *Archives of Virology*, *160*(4), 1113–1118.

Annan, A., Baldwin, H. J., Corman, V. M., Klose, S. M., Owusu, M., Nkrumah, E. E., Drexler, J. F. (2013). Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe. *Emerging Infectious Diseases*, *19*(3), 456–459.

Anthony SJ, Ojeda-Flores R, Rico-Chávez O, Navarrete-Macias I, Zambrana-Torrelio CM, Rostal MK, Epstein JH, Tipps T, Liang E, Sanchez-Leon M, Sotomayor-Bonilla J, Aguirre AA, Ávila-Flores R, Medellín RA, Goldstein T, Suzán G, Daszak P, Lipkin WI. (2013). Coronaviruses in bats from Mexico. The Journal of General Virology, 94(Pt 5), 1028

Anthony, Simon J., Johnson, C. K., Greig, D. J., Kramer, S., Che, X., Wells, H., L. H. Allison, O. J. Damien, D. W. Nathan, D. Peter, K. William,W. I. Lipkin, S. M. Morse, PREDICT Consortium, A. K. M.Jonna, Goldstein, T. (2017). Global patterns in coronavirus diversity. Virus Evolution, 3(1), 1–15.

Bininda-Emonds, O. R. P., Cardillo, M., Jones, K. E., MacPhee, R. D. E., Beck, R. M. D., Grenyer, R.,Price S.A, Vos R. A,. Gittleman J. L, Purvis, A. (2007). The delayed rise of present-day mammals. Nature, 446(7135), 507–512.

Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V., & Schountz, T. (2006). Bats: Important reservoir hosts of emerging viruses. Clinical Microbiology Reviews, 19(3), 531–545.

Carrington, C. V. F., Foster, J. E., Zhu, H. C., Zhang, J. X., Smith, G. J. D., Thompson, N., Auguste A.J, Ramkissoon V., Adesiyun A.A., Guan, Y. (2008). Detection and phylogenetic analysis of group 1 coronaviruses in South American bats. Emerging Infectious Diseases, 14(12), 1890–1893.

Chen, L., Liu, B., Yang, J., & Jin, Q. (2014). DBatVir: The database of bat-associated viruses. Database, 2014, 1–7.

Chu, D. K. W., Leung, C. Y. H., Gilbert, M., Joyner, P. H., Ng, E. M., Tse, T. M., Poon, L. L. M. (2011). Avian Coronavirus in Wild Aquatic Birds. *Journal of Virology*, *85*(23), 12815–12820.

Consortium, P. (2014). Reducing pandemic risk, promoting global health. One Health Institute, Davis, CA: University of California, Davis.

Constantine, D. G. (2003). Geographic translocation of bats: known and potential problems. Emerging Infectious Diseases, 9(1), 17.

Corman, V. M, Muth, D., Niemeyer, D., & Drosten, C. (2018). Hosts and Sources of Endemic Human Coronaviruses. In Advances in Virus Research (Vol. 100, pp. 163–188).

Corman, V. M., Baldwin, H. J., Tateno, A. F., Zerbinati, R. M., Annan, A., Owusu, M., Nkrumah, E. E, Maganga G. D., Oppong S., Adu-Sarkodia Y., Vallo P., Filho L.V.R.F.S, Leroy E. M., Thiel V., van der Hoek L., Poon L. L. M, Tschapka M., Drosten, C., Drexler, J. F. (2015). Evidence for an Ancestral Association of Human Coronavirus 229E with Bats. Journal of Virology, 89(23), 11858–11870.

Corman, V. M., Ithete, N. L., Richards, L. R., Schoeman, M. C., Preiser, W., Drosten, C., & Drexler, J. F.(2014). Rooting the Phylogenetic Tree of Middle East Respiratory Syndrome Coronavirus by Characterization of a Conspecific Virus from an African Bat. Journal of Virology, 88(19), 11297–11303.

Corman, V. M., Kallies, R., Philipps, H., Gopner, G., Muller, M. A., Eckerle, I., Brünink S., Drosten C.,Drexler, J. F. (2014). Characterization of a Novel Betacoronavirus Related to Middle East Respiratory Syndrome Coronavirus in European Hedgehogs. Journal of Virology, 88(1), 717–724.

Corman, V.M, Rasche, A., Diallo, T. D., Cottontail, V. M., Stöcker, A., Souza, B. F. de C. D.,Corrêa, J. I., Carneiro ,A. J. B, Franke, C.R., Nagy, M., Metz, M., Knörnschild, M.,Kalko, E. K. V., Ghanem S.J, Morales, K. D. S., Salsamendi, E., Spínola, M., Herrler, G., Voigt, C. C., Tschapka, M., Drosten, C., Drexler, J. F. (2013). Highly diversified coronaviruses in neotropical bats. Journal of General Virology, 94(PART9), 1984–1994

De Groot, R. J., Baker, S. C., Baric, R., Enjuanes, L., Gorbalenya, A., Holmes, K. V., Perlman S., Poon L., Rottier P. J. M.,Talbot, P. J, Woo P.C.Y, Ziebuhr J.,(2012). Family Coronaviridae, p 806 – 820. In King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed), Virus taxonomy: classification and nomenclature of viruses. Ninth report of the International Committee on Taxonomy of Viruses. Academic Press, London, United Kingdom

de Haan, C. A., Vennema, H., & Rottier, P. J. (2000). Assembly of the coronavirus envelope: homotypic interactions between the M proteins. Journal of Virology, 74(11), 4967–4978.

de Souza Luna, L. K., Heiser, V., Regamey, N., Panning, M., Drexler, J. F., Mulangu, S.,Drosten, C. (2007). Generic detection of coronaviruses and differentiation at the prototype strain level by reverse transcription-PCR and nonfluorescent low-density microarray. *Journal of Clinical Microbiology*, *45*(3), 1049–1052.

Drexler, J. F., Corman, V. M., & Drosten, C. (2014). Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. Antiviral Research, 101(1), 45–56.

Drexler, J. F., Gloza-rausch, F., Corman, V. M., Muth, D., Goettsche, M., Seebens, A. Niedrig M., Pfefferle

S., Yordanov S., Zhelyakov L., Hermanns U., Vallo P., Lukashev A., Muller M. A., Deng H., Herrler G., Drosten, C. (2010). coronavirus in European bats and classification of coronaviruses based on partialRNA-dependent RNA polymerase gene sequences. J. Virol. 84, 11336– 11349

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian P hylogenetics with BEAUti and the BEAST 1 . 7. 29(8), 1969–1973.

Falcón, A., Vázquez-Morón, S., Casas, I., Aznar, C., Ruiz, G., Pozo, F., Chevarría, J. E. (2011). Detection of alpha and betacoronaviruses in multiple Iberian bat species. *Archives of Virology*, *156*(10), 1883–1890.

Fehr, A. R., & Perlman, S. (2015). Coronaviruses: An overview of their replication and pathogenesis. In *Coronaviruses: Methods and Protocols*.

Gallagher, T. M., & Buchmeier, M. J. (2001). Coronavirus spike proteins in viral entry and pathogenesis. *Virology*, *279*(2), 371–374.

Góes, L. G. B., Campos, A. C. de A., Carvalho, C. De, Ambar, G., Queiroz, L. H., Cruz-Neto, A. P.,Munir M.,Durigon, E. L. (2016). Genetic diversity of bats coronaviruses in the Atlantic Forest hotspot biome, Brazil. Infection, Genetics and Evolution, 44, 510–513.

Gong, L., Li, J., Zhou, Q., Xu, Z., Chen, L., Zhang, Y., Xue C., Wen Z., Cao, Y. (2017). A new bat-HKU2–like coronavirus in swine, China, 2017. Emerging Infectious Diseases, 23(9), 1607–1609.

Gorbalenya, A. E., Snijder, E. J., & Ziebuhr, J. (2015). Virus-encoded proteinases and proteolytic processing in the Nidovirales. Journal of General Virology, 81(4), 853–879.

Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Cheung, C. L.,. Lou S. W., Zhang L.J., Guan, Y. J., Butt K.M., Wong K. W., Lim W., Shortridge K. F., Yuen K.Y., Peiris J.S., Poom L.L.M (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science, 302(5643), 276–278.

Hon, C., Lam, T., Shi, Z., Drummond, A. J., Yip, C., Zeng, F., Lam P-Y., Leung, F. C. (2008). Evidence of the Recombinant Origin of a Bat Severe Acute Respiratory Syndrome (SARS) -Like Coronavirus and Its Implications on the Direct Ancestor of SARS Coronavirus . 82(4), 1819–1826.

Hu, H., Jung, K., Wang, Q., Saif, L. J., & Vlasova, A. N. (2018). Development of a one-step RT-PCR assay for detection of pancoronaviruses (α-, β-, γ-, and δ-coronaviruses) using newly designed degenerate primers for porcine and avian 'fecal samples. *Journal of Virological Methods*, *256*(January), 116–122.

Kaslow, R. A., Stanberry, L. R., & Le Duc, J. W. (2014). Viral infections of humans: Epidemiology and

control. *Viral Infections of Humans: Epidemiology and Control*, 1–1215.

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). *MAFFT : a novel method for rapid multiple sequence alignment based on fast Fourier transform*. *30*(14), 3059–3066.

Katoh, K., & Standley, D. M. (2013). *MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track*. *30*(4), 772–780.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*, 845. Retrieved from

Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong S., Urbani C., Comer J.A., Lim W., Rollin P.E., Dowell S.F., Ling A.E., Humphrey C.D., Shieh W.J., Guarner J., Paddock C.D., Rota P., Fields B. DeRisi J., Yang J.Y., Cox N., Hughes J.M., LeDuc J.W., Bellini W.J., Anderson, L. J. (2003). A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. New England Journal of Medicine, 348(20), 1953–1966.

Lacroix, A., Duong, V., Hul, V., San, S., Davun, H., Omaliss, K.,Buchy, P. (2017). Genetic diversity of coronaviruses in bats in Lao PDR and Cambodia. *Infection, Genetics and Evolution*, *48*, 10–18.

Lai, & C., M. (2007). Coronaviridae. In *Fields Virology* (pp. 1305–1318).

Lai, M. M., and D. Cavanagh. (1997). The molecular biology of coronaviruses. Adv. Virus Res. 48:1–100.

Lau, S. K. P., Lee, P., Tsang, A. K. L., Yip, C. C. Y., Tse, H., Lee, R. A., So L.Y., Lau Y.L., Chan K.H., Woo P.C.YYuen, K.-Y. (2011). Molecular Epidemiology of Human Coronavirus OC43 Reveals Evolution of Different Genotypes over Time and Recent Emergence of a Novel Genotype due to Natural Recombination. Journal of Virology, 85(21), 11325–11337.

Lau, Susanna K P, Li, K. S. M., Huang, Y., Shek, C., Tse, H., Wang, M., Choi G.K.Y, Xu H., Lam C.S.F., Guo R., Chan K.H., Zheng B.J., Woo p.C.Y., Yuen K.Y.,(2010). Ecoepidemiology and Complete Genome Comparison of Different Strains of Severe Acute Respiratory Syndrome-Related Rhinolophus Bat Coronavirus in China Reveal Bats as a Reservoir for Acute , Self-Limiting Infection That Allows Recombination Events Journal of Virology Feb 2010, 84 (6) 2808-2819

Lau, Susanna K P, Woo, P. C. Y., Li, K. S. M., Huang, Y., Wang, M., Lam, C. S. F., Xu H., Guo R., Chan K.H., Zheng B.J., Zheng, B. (2007). Complete genome sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome. Virology, 367(2), 428–439.

Lelli, D., Papetti, A., Sabelli, C., Rosti, E., Moreno, A., & Boniotti, M. B. (2013). Detection of coronaviruses

in bats of various species in Italy. *Viruses*, *5*(11), 2679–2689.

Lim, Y., Ng, Y., Tam, J., & Liu, D. (2016). Human Coronaviruses: A Review of Virus–Host Interactions. Diseases, 4(4), 26.

López-Aguirre, C., Hand, S. J., Laffan, S. W., & Archer, M. (2018). Phylogenetic diversity, types of endemism and the evolutionary history of New World bats. Ecography, 41(12), 1955–1966.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4 : Detection and analysis of recombination patterns in virus genomes. 1–5.

Masters, P. S. (2006). The Molecular Biology of Coronaviruses. Advances in Virus Research, 65(06), 193–292.

Memish, Z. A., Mishra, N., Olival, K. J., Fagbo, S. F., Kapoor, V., Epstein, J. H., Alhakeem R., Durosinloun A., Al Asmari M., Islam A., Kapoor A., Briese T., Daszak P., Al Rabeeah A.A., Lipkin, W. I. (2013). Middle East respiratory syndrome coronavirus in Bats, Saudi Arabia. Emerging Infectious Diseases, 19(11), 1819–1823.

Moratelli, R., & Calisher, C. H. (2015). Bats and zoonotic viruses: Can we confidently link bats with emerging deadly viruses? Memorias Do Instituto Oswaldo Cruz, 110(1), 1–22.

Moreira-Soto, A., Taylor-Castillo, L., Vargas-Vargas, N., Rodríguez-Herrera, B., Jiménez, C., & Corrales-Aguilar, E. (2015). Neotropical Bats from Costa Rica harbour Diverse Coronaviruses. *Zoonoses and Public Health*, *62*(7), 501–505.

Neuman, B. W., Joseph, J. S., Saikatendu, K. S., Serrano, P., Chatterjee, A., Johnson, M. A., Liao L., Klaus J.P.,Yates L.R 3rd., Wüthrich K., Stevens R.C., Buchmeier M.J., Kuhn, P. (2008). Proteomics Analysis Unravels the Functional Repertoire of Coronavirus Nonstructural Protein 3. Journal of Virology, 82(11), 5279–5294.

Nieva, J. L., & Carrasco, L. (2015). Viroporins: Structures and functions beyond cell membrane permeabilization. *Viruses*, *7*(10), 5169–5171.

Peixoto, F. P., Henrique, P., Braga, P., & Mendes, P. (2018). A synthesis of ecological and evolutionary determinants of bat diversity across spatial scales. *BMC Ecology*, 1–14.

Pfleiderer, M., Routledge, E., Herrler, G., & Siddell, S. G. (1991). High level transient expression of the murine coronavirus haemagglutinin-esterase. *Journal of General Virology*, *72*(6), 1309–1315.

Pyrc, K., Dijkman, R., Deng, L., Jebbink, M. F., Ross, H. A., Berkhout, B., & van der Hoek, L. (2006). Mosaic Structure of Human Coronavirus NL63, One Thousand Years of Evolution. *Journal of Molecular*

Biology, *364*(5), 964–973.

Quan, P., Firth, C., & Street, C. (2010). Identification of a Severe Acute Respiratory Syndrome Coronavirus-. *MBio*, *1*(4), 1–9.

Rex, K., Kelm, D. H., Wiesner, K., Kunz, T. H., & Voigt, C. C. (2008). Species richness and structure of three phyllostomid bat assemblages. *Bat Research News*, *47*, 138–139.

Ro**nquist, F., & Huelsenbeck, J. P.** (2003). *MrBayes 3 : Bayesian phylogenetic inference under mixed models*. *19*(12), 1572–1574.

Rottier, P. J. M. (1995). The Coronavirus Membrane Glycoprotein. In S. G. Siddell (Ed.), *The Coronaviridae* (pp. 115–139).

Sabir, J. S. M., Ahmed, M. M. M., Li, L., Shen, Y., Abo-aba, S. E. M., Qureshi, M. I., Abu-Zeid M., Zhang Y., Khiyami M.A., Alharbi N.S., Hajarah N.H., Sabir M.J., Mutwakil M.H., Kabli S.A., Alsulaimany F.A., Obaid A.Y., Zhou B., Smith D.K., Holmes E.C., Zhu H., Guan Y., (2016). Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. 351(6268), 81–85.

Saif, L. J. (2004). Animal coronaviruses: what can they teach us about the severe acute respiratory syndrome? *Revue Scientifique et Technique-Office International Des Épizooties*, *23*(2), 643–660.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463 LP – 5467.

Sawicki, S. G., Sawicki, D. L., & Siddell, S. G. (2007). A Contemporary View of Coronavirus Transcription. *Journal of Virology*, *81*(1), 20–29.

Schultze, B., Wahn, K., Klenk, H.-D., & Herrler, G. (1991). Isolated HE-protein from hemagglutinating encephalomyelitis virus and bovine coronavirus has receptor-destroying and receptor-binding activity. *Virology*, *180*(1), 221–228.

Simmonds, P. (2012). SSE: A nucleotide and amino acid sequence analysis platform. *BMC Research Notes*, *5*(January).

Simmons, N. B. (2005). Order chiroptera. In *Mammal species of the world: a taxonomic and geographic reference* (pp. 312–529). Baltimore, MD: Johns Hopkins University Press.

Sola, I., Almazán, F., Zúñiga, S., & Enjuanes, L. (2015). Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annual Review of Virology*, *2*(1), 265–288.

Sola, I., Mateos-Gomez, P. A., Almazan, F., Zuñiga, S., & Enjuanes, L. (2011). RNA-RNA and RNA-protein

interactions in coronavirus replication and transcription. *RNA Biology*, *8*(2).

Solari, S. (2016). Pteronotus paraguanensis. *The IUCN Red List of Threatened Species 2016: E.T136610A21987754*, *8235*.

Stadler, K., Masignani, V., Eickmann, M., & Becker, S. (2003). SARS — Beginning to understand a new virus. Nature Reviews Microbiology 1, 209–218

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–2729.

Tao, Y., Shi, M., Chommanard, C., Queen, K., & Zhang, J. (2017). crossm Surveillance of Bat Coronaviruses in. 91(5), 1–16.

Tao, Y., Tang, K., Shi, M., Conrardy, C., Li, K. S. M., Lau, S. K. P., Anderson L.J., Tong, S. (2012). Genomic characterization of seven distinct bat coronaviruses in Kenya. Virus Research, 167(1), 67–73.

Teeling, E. C., Springer, M. S., Madsen, O., Bates, P., O'Brien, S. J., & Murphy, W. J. (2005). A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science*, *307*(5709), 580–584.

Vijgen, L., Moës, E., Keyaerts, E., Li, S., & Van Ranst, M. (2008). A Pancoronavirus RT-PCR Assay for Detection of All Known Coronaviruses BT - SARS- and Other Coronaviruses: Laboratory Protocols (D. Cavanagh, ed.).

Vlasak, R., Luytjes, W., Leider, J., Spaan, W., & Palese1, P. (1988). The E3 Protein of Bovine Coronavirus Is a Receptor-Destroying Enzyme with Acetylesterase Activity. *Journal of Virology*, *62*(12), 4686–4690.

Wacharapluesadee, S., Duengkae, P., Rodpan, A., Kaewpom, T., Maneeorn, P., Kanchanasaka, B., Hemachudha, T. (2015). Diversity of coronavirus in bats from Eastern Thailand. *Virology Journal*, *12*(1), 57.

Watanabe, S., Masangkay, J. S., Nagata, N., Morikawa, S., Mizutani, T., Fukushi, S., Akashi, H. (2010). Bat coronaviruses and experimental infection of bats, the Philippines. *Emerging Infectious Diseases*, *16*(8), 1217–1223.

Weiss, S. R., & Navas-Martin, S. (2005). Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus. *Microbiology and Molecular Biology Reviews*, *69*(4), 635–664.

Wong, A. C. P., Li, X., Lau, S. K. P., & Woo, P. C. Y. (2019). Global epidemiology of bat coronaviruses. *Viruses*, *11*(2), 1–17.

Woo, P. C. Y., Lau, S. K. P., Lam, C. S. F., Tsang, A. K. L., Hui, S.-W., Fan, R. Y. Y., Martelli P., Yuen, K.-Y.(2014). Discovery of a Novel Bottlenose Dolphin Coronavirus Reveals a Distinct Species of Marine Mammal Coronavirus in Gammacoronavirus. Journal of Virology, 88(2), 1318–1331.

Woo, Patrick C. Y., Lau, S. K. P., Huang, Y., & Yuen, K.-Y. (2009). Coronavirus Diversity, Phylogeny and Interspecies Jumping. *Experimental Biology and Medicine*, *234*(10), 1117–1127.

Woo, Patrick C.Y., Lau, S. K. P., Li, K. S. M., Poon, R. W. S., Wong, B. H. L., Tsoi, H. wah, Yip B.C.K., Huang Y., Chan K.Y., Yuen,K.Y. (2006). Molecular diversity of coronaviruses in bats. Virology, 351(1), 180–187.

Woo, Patrick C Y, Lau, S. K. P., Lam, C. S. F., Lau, C. C. Y., Tsang, A. K. L., Lau, J. H. N., Bai R., Teng J.L.L., Tsang C.C.C., Wang M., Zheng B.J., Chan K.H., Yuen, K.-Y. (2012). Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene source of Gammacoronavirus and Deltacoronavirus. Journal of Virology, 86(7), 3995 LP – 4008.

Xu, D., Zhang, Z., Chu, F., Li, Y., Jin, L., Zhang, L., Gao G.F., Wang, F. S. (2004). Genetic Variation of SARS Coronavirus in Beijing Hospital. Emerging Infectious Diseases, 10(5), 789–794.

Zhao, L., Jha, B. K., Wu, A., Elliott, R., Ziebuhr, J., Gorbalenya, A. E.,Weiss, S. R. (2012). Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology. Cell Host & Microbe, 11(6), 607–616.

Zhou, P., Fan, H., Lan, T., Yang, X., Shi, W., Zhang, W., Zhu Y., Zhang Y.W., Xie Q.M., Mani S., Zheng X.S., Li B., Li J.M., Guo H., Pei G.Q., An X.P., Chen J.W., Zhou L., Mai K.J., Wu Z.X., Li D., Anderson D.E., Zhang L.B., Li S.Y., Mi Z.Q., He T.T., Cong F., Guo P.J., Huang R., Luo Y., Liu X.L., Chen J., huang Y., Sun Q., Zhang X.L.L., Wnag Y.Y., Xing S.Z., Chen Y.S., Sun Y., Li J., Daszak P., Wang L.F., Shi Z.L., Tong Y.G., Ma J.Y (2018). HKU2-related coronavirus of bat origin. Nature 556 255-258.

6. Supplementary section

S1 – Table with the accession numbers of the 229E and NL63 sequences used for the phylogenies

Name	Accession number (GenBank)
229E-related bat coronavirus isolate BtCoV/KW2E-F56/Hip_cf._rub/GHA/2011	KT253271.1
229E-related bat coronavirus isolate BtCoV/AT1A-F1/Hip_aba/GHA/2010	KT253272.1
Human coronavirus 229E strain 229E/human/USA/933-50/1993	KF514430.1
Human coronavirus 229E isolate J0304	JX503061.1
Human coronavirus 229E isolate 0349	JX503060.1
Human coronavirus 229E strain 229E/UF-1/2016	KY996417.1
Human coronavirus 229E strain SC9773	KY369914.1
Human coronavirus 229E strain SC2282	KY684760.1
Human coronavirus 229E strain 229E/Haiti-1/2016	MF542265.1
Human coronavirus 229E strain SC677	KY369909.2
Human coronavirus 229E strain SC399	KY674914.1
Human coronavirus 229E strain SC1073	KY369913.1
Human coronavirus 229E strain SC1143	KY369910.1
Human coronavirus 229E strain SC1212	KY369911.1
Human coronavirus 229E strain N08-434B	KY674919.1
Human coronavirus 229E strain SC9731	KY369912.1
Human coronavirus 229E isolate HCoV-229E/BN1/GER/2015	KU291448.1
229E-related bat coronavirus strain BtKY229E-1	KY073747.1
Alpaca respiratory coronavirus isolate CA08-1/2008	JQ410000.1
229E-related bat coronavirus isolate BtCoV/FO1A-F2/Hip_aba/GHA/2010	KT253270.1
Camel alphacoronavirus Camel229E isolate Camel229E-CoV/JC50/KSA/2014	KT253324.1
Camel alphacoronavirus Camel229E isolate Camel229E-CoV/JC49/KSA/2014	KT253325.1
Camel_alphacoronavirus_Camel229E_isolate_Camel229E-CoV/JC52/KSA/2014 extraction	KT253326.1
Camel alphacoronavirus isolate camel/Jeddah/Jd29/2014	KT368892.1
Camel alphacoronavirus isolate camel/Taif/T91(a)/2015	KT368913.1
Camel alphacoronavirus isolate camel/Jeddah/O23(a)/2014	KT368895.1
Camel alphacoronavirus Abu Dhabi B38	MF593473.1
Camel alphacoronavirus isolate camel/Riyadh/Ry141/2015	NC_028752.1
Camel alphacoronavirus Camel229E isolate Camel229E-CoV/AC04/KSA/2014	KT253327.1
Camel alphacoronavirus Camel229E isolate Camel229E-CoV/KCSP1/KEN/2015	KU291449.1
229E-related bat coronavirus isolate BtCoV/KW2E-F151/Hip_cf._rub/GHA/2011	KT253269.1
Human Coronavirus NL63	NC_005831.2
Human coronavirus NL63 isolate NL63/RPTEC/2004	JX504050.1
Human coronavirus NL63 strain NL63/DEN/2005/193	JQ765568.1
Human coronavirus NL63 strain N07-468B_176X	KY554971.1
Human coronavirus NL63 strain NL63/human/USA/012-31/2001	KF530105.1
Human coronavirus NL63 strain NL63/human/USA/904-20/1990	KF530104.1

Human coronavirus NL63 strain NL63/human/USA/891-6/1989	KF530108.1
Human coronavirus NL63 isolate Amsterdam 057	DQ445911.1
Human coronavirus NL63 strain SC2940	KY983586.1
Human coronavirus NL63 isolate CN0601/14	MG772808.1
Human coronavirus NL63 isolate NL63/UF-1/2015	KT381875.1
Human coronavirus NL63 strain NL63/DEN/2009/20	JQ765567.1
Human coronavirus NL63 strain N07-324B_182X	KY554970.1
Human coronavirus NL63 strain N06-1144B	KY554967.1
Human coronavirus NL63 strain N07-185B	KY554968.1
Human coronavirus NL63 strain NL63/DEN/2008/16	JQ765566.1
Human coronavirus NL63 strain NL63/DEN/2005/232	JQ765569.1
Human coronavirus NL63 strain NL63/DEN/2009/9	JQ765563.1
Human coronavirus NL63 strain NL63/human/USA/838-9/1983	KF530110.1
Human coronavirus NL63 isolate Kilifi_HH_5402_20-May-2010	MG428704.1
Human coronavirus NL63 isolate CBJ 037	JX104161.1
Human coronavirus NL63 isolate CBJ123	JX524171.1
Human coronavirus NL63 isolate HCoV NL63/Haiti-1/2015	KT266906.1
Human coronavirus NL63 isolate Kilifi_HH_3807_11-May-2010	MG428702.1
NL63-related bat coronavirus strain BtKYNL63-9a	NC_032107.1
NL63-related bat coronavirus strain BtKYNL63-9b	KY073745.1